

Luminous Red Galaxy Halo Density Field Reconstruction and Application to Large Scale Structure Measurements

Beth A. Reid

Department of Physics

Department of Astrophysical Sciences

Princeton University, Princeton, NJ 08544

`breid@princeton.edu`

David N. Spergel

Department of Astrophysical Sciences

Princeton Center for Theoretical Sciences

Princeton University, Princeton, NJ 08544

`dns@astro.princeton.edu`

and

Paul Bode

Department of Astrophysical Sciences

Princeton University, Princeton, NJ 08544

`bode@astro.princeton.edu`

ABSTRACT

The complex relationship between the galaxy density field and the underlying matter field limits our ability to extract cosmological constraints from galaxy redshift surveys. Our approach is to use halos rather than galaxies to trace the underlying mass distribution. We identify Fingers-of-God (FOGs) and replace multiple galaxies in each FOG with a single halo object. This removes the nonlinear contributions of satellite galaxies, the one-halo term. We test our method on a large set of high-fidelity mock SDSS Luminous Red Galaxy (LRG) catalogs. We find that the aggressive FOG compression algorithm

adopted in the LRG $P(k)$ analysis of Tegmark et al. (2006) leads to a $\sim 10\%$ correction to the underlying matter power spectrum at $k = 0.1 \ h \text{ Mpc}^{-1}$ and $\sim 40\%$ correction at $k = 0.2 \ h \text{ Mpc}^{-1}$, thereby compromising the cosmological constraints. In contrast, the power spectrum of our reconstructed halo density field deviates from the underlying matter power spectrum at the $\leq 1\%$ level for $k \leq 0.1 \ h \text{ Mpc}^{-1}$ and $\leq 4\%$ at $k = 0.2 \ h \text{ Mpc}^{-1}$. The reconstructed halo density field also removes the bias in the measurement of the redshift space distortion parameter β induced by the FOG smearing of the linear redshift space distortions.

Subject headings: cosmology: observations — galaxies: halos — galaxies: statistics — galaxies:elliptical and lenticular, cD

1. Introduction

Galaxy power spectra are an essential element in testing our cosmological models and studying the initial conditions of our universe. As surveys continue to grow in size, the statistical error bars on the galaxy power spectra continue to shrink. However, our ability to constrain cosmological models is limited by the systematics of relating the galaxy distribution to the underlying matter field. For example, through a comparison of galaxy samples from 2dFGRS and SDSS, Sanchez & Cole (2007) find that differing scale-dependent biasing between red and blue galaxies cause the cosmological constraints derived from the samples to differ at the $\sim 2\sigma$ level. These systematic problems are even more severe in the SDSS Luminous Red Galaxy (LRG) sample. With its large effective volume (Tegmark et al. 2006), this sample potentially provides the tightest constraints on the shape of the linear power spectrum. However, because of the low number density of the LRGs and their high bias, the LRG power spectrum potentially requires a large nonlinear correction. As a consequence, several authors have encountered problems when combining the LRG power spectrum results of Tegmark et al. (2006) with CMB data sets. Dunkley et al. (2008) find that when combining with the 5 year WMAP results, the best fit Ω_m varies systematically with the maximum k from the LRG $P(k)$ included in the analysis. When Verde & Peiris (2008) reconstruct the primordial power spectrum with minimal assumptions about its shape, they find that the LRG sample has less statistical power than the SDSS MAIN or 2dFGRS, despite its larger effective volume. Since the authors restrict themselves to $k \leq 0.1$, they are unable to constrain both the nonlinear correction and the primordial power spectrum. The difficulties in both analyses stem from the degeneracy between the power spectrum shape and the potentially large nonlinear correction amplitude. Fig. 1 shows that the best

fit Tegmark et al. (2006) nonlinear correction is above the statistical error in each k -bin for $k \gtrsim 0.07$, while $k \lesssim 0.1-0.15$ is typically used in cosmological parameter analyses. Moreover, the k -dependence of this large nonlinear correction must be accurate to avoid introducing systematic errors. The solution we propose in this paper is to first estimate a halo density field from the LRG density field. The resulting field is nearly linearly biased with respect to the dark matter for $k \leq 0.2$. The dashed curve in Fig. 1 shows that the nonlinear correction between the dark matter and linear spectrum is much smaller than the correction applied in the Tegmark et al. (2006) LRG analysis; it is also much easier to calibrate using N -body simulations.

The Halo Occupation Distribution (HOD) model (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002) allows us to describe the link between galaxies and dark matter, and provides a framework in which to analyze non-linearities contributing to the power spectrum. Pairs of galaxies can be separated as one- or two-halo, i.e., occupying the same or distinct dark matter halos. As discussed in Seljak (2000) and Schulz & White (2006), scale dependence of the bias between galaxies and dark matter arises primarily because the one-halo galaxy term is enhanced relative to the dark matter more than the two-halo term. To combat this effect, Huff et al. (2007) introduce a configuration space band-power estimator which confines the one-halo contribution to $\sim 2 - 3 h^{-1}$ Mpc. Hamann et al. (2008) find that for measured galaxy power spectra (SDSS MAIN, 2dFGRS, and SDSS LRG samples), the nonlinearity can be sufficiently modeled as excess shot noise.

In extracting the underlying linear matter power spectrum from large galaxy redshift surveys, there are three main approaches to disentangling the effects of nonlinear evolution in the matter field, nonlinear redshift space distortions, and nonlinear galaxy biasing. One can model the nonlinearities with a phenomenological fitting formula and marginalize over its free parameters when fitting cosmological parameters; this technique has been employed in analyses of SDSS LRGs (Tegmark et al. 2006) as well as in 2dFGRS (Cole et al. 2005). The work of Yoo et al. (2008) instead starts with the HOD to predict the nonlinear modifications to large-scale clustering statistics. In principal, cosmological and HOD parameters can be fit simultaneously. In this forward approach, the measured large-scale clustering is compared with predicted clustering given the HOD parameters. The method put forth in this work is an *inverse* method: we make use of phase information in the galaxy density field to estimate a halo density field. The power spectrum of this field is nearly linearly biased with respect to the underlying nonlinear matter density field. A future extension of this work is to also use the phase information to reconstruct the linear density field (Eisenstein et al. 2007a). This combined approach would “undo” two sources of nonlinearity: galaxy bias and nonlinear

dynamics.

In this paper we introduce our halo density field reconstruction method and use N -body simulations to test the method. We first review previous approaches to the analysis of the LRG $P(k)$ in § 2. In § 3 and 4 we apply the CiC method of Reid & Spergel (2008) to produce high fidelity mock catalogs from 42 N -body simulations for three LRG redshift subsamples: NEAR, MID, and FAR. § 5 describes our method for reconstructing a halo density field. In § 6 we present the nonlinear dark matter and mock catalog power spectra and covariance matrices. We analyze in detail the effects of the Finger-of-God (FOG) compression algorithm employed in the Tegmark et al. (2006) analysis of the LRG power spectrum. We also present results for the angle-averaged redshift space spectrum, which is measured in Percival et al. (2007). We demonstrate that the reconstructed halo density field is linearly biased with respect to the dark matter out to $k = 0.2$ at the 2% level for the MID and FAR subsamples and at the 4% level for the NEAR sample. For $k \leq 0.1$, the traditional regime used for cosmological parameter estimation, the discrepancy is $\leq 1\%$. We also show that the beat-coupling model of Hamilton et al. (2006) fits both the dark matter and mock catalog covariance matrices. Finally we examine the structure of the redshift space distortions as a function of k . Throughout this paper we adopt the cosmological parameters recommended from the latest WMAP5 analysis (Komatsu et al. 2008): $(\Omega_m, \Omega_b, \Omega_\Lambda, n_s, \sigma_8, h) = (0.2792, 0.0462, 0.7208, 0.960, 0.817, 0.701)$.

2. Background

2.1. Summary of Previous Analyses of the SDSS LRG Power Spectrum

Two power spectrum analyses were simultaneously published for the fifth data release of the SDSS. Percival et al. (2007) compute the monopole (i.e., angle-averaged) redshift space power spectrum of the combined sample of MAIN and LRG galaxies. The method extends the FKP method (Feldman et al. 1994) to remove differential bias from the recovered spectrum, and optimally weights the galaxies according to their expected bias, which is determined from their luminosity. This approach leads Percival et al. (2007) to use a different selection function for the LRGs than adopted for our analysis. We do not model the luminosities of the galaxies, and so precise evaluation of their method is not possible here. However, we do compute the monopole redshift power spectrum of our mock catalogs. This should compare most closely to the Percival et al. (2007) method, though the weighting of the galaxies according to their luminosity will certainly alter the relative amplitudes of the

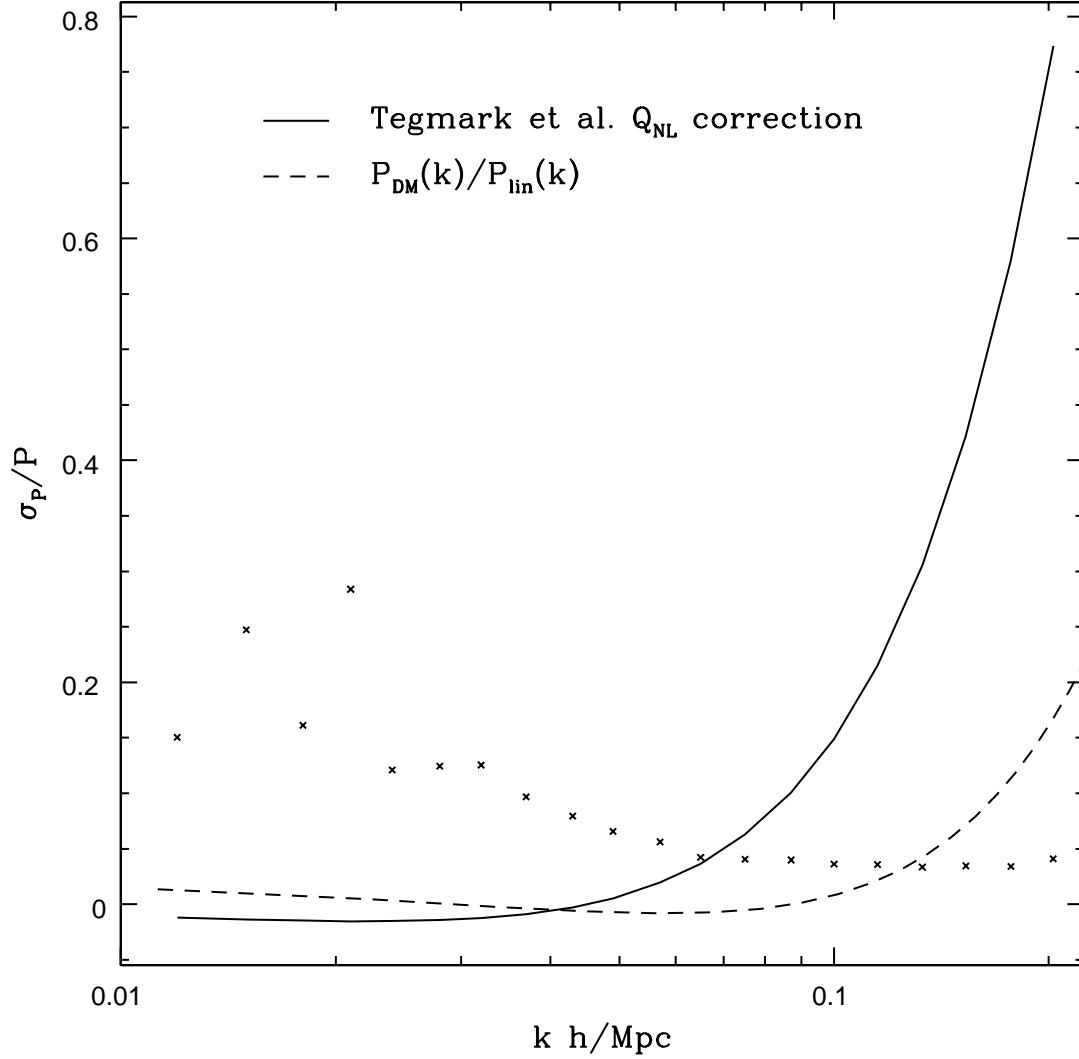


Fig. 1.— Points show the error bars divided by the bandpower for the LRG power spectrum published in Tegmark et al. (2006). The solid curve shows $(1 + Q_{NL}k^2)/(1 + 1.4k) - 1$, the fractional amplitude of the nonlinear Q_{NL} correction to the linear power spectrum for $Q_{NL} = 31$, the best fit value in Tegmark et al. (2006). For comparison, the dashed curve shows the smooth nonlinear correction fit to the dark matter power spectrum of Fig. 5. This is the size of the correction for our reconstructed halo density field.

one- and two-halo contributions to the power spectrum.

Tegmark et al. (2006) focus their analysis entirely on the LRGs. They examine three redshift subsamples: NEAR ($0.155 < z < 0.300$), MID ($0.300 < z < 0.380$), and FAR ($0.380 < z < 0.474$). We model each of these samples separately in our analysis. Tegmark et al. (2006) use the PKL method (Tegmark et al. 2004) to estimate the real space galaxy-galaxy, galaxy-velocity, and velocity-velocity power spectra. In linear theory the three spectra are related by (Eqns. 2 and 3 in Tegmark et al. 2006):

$$\begin{aligned} P_{gv}(k) &= \beta r_{gv} P_{gg}(k) \\ P_{vv}(k) &= \beta^2 P_{gg}(k). \end{aligned} \tag{1}$$

β relates the amplitude of galaxy and velocity field fluctuations in linear theory (Eqn. 35) and r_{gv} is the dimensionless correlation coefficient between the velocity field and gravitational acceleration field. The final estimate of the real space LRG power spectrum is a linear combination of all three, though it is dominated by the scaled redshift space monopole power spectrum. Their method assumes that the parameters relating the galaxy and velocity power spectra, β and r_{gv} , are scale independent, and they only use data with $k < 0.09 \, h \, \text{Mpc}^{-1}$ to estimate these factors. To make this assumption more accurate for the LRG density field, Tegmark et al. (2006) compress FOGs before estimating the power spectrum. The algorithm is detailed in Tegmark et al. (2004), but to our knowledge has not been extensively tested on accurate LRG mock catalogs. The algorithm finds groups of LRGs by a Friends-of-Friends (FoF) algorithm (Frenk et al. 1988) and then isotropizes them by scaling all radial separations from the group center. Two objects are considered friends when

$$\left[\left(\frac{r_{\parallel}}{10} \right)^2 + r_{\perp}^2 \right]^{1/2} \leq \left[\frac{4}{3} \pi \bar{n} (1 + \delta_c) + r_{\perp, \max}^{-3} \right]^{-1/3} \tag{3}$$

with $\delta_c = 200$ and $r_{\perp, \max} = 5 \, h^{-1} \, \text{Mpc}$; the RHS of Eqn. 3 is $2.3 \, h^{-1} \, \text{Mpc}$ for the SDSS LRG NEAR and MID subsample number densities, and $2.8 \, h^{-1} \, \text{Mpc}$ for the FAR sample. For comparison, the virial radius of a $10^{14} h^{-1} M_{\odot}$ halo is $\sim 1 \, h^{-1} \, \text{Mpc}$ (~ 2 for a $10^{15} h^{-1} M_{\odot}$ halo). Since the line-of-sight (LOS) distance is compressed in the LHS of Eqn. 3, the algorithm will compress objects well beyond the virial radius of the typical host halo. After examination of the NEAR, MID, and FAR subsamples, Tegmark et al. (2006) conclude that the expected decrease in bias and increase in clustering amplitude with redshift cancel within the observational errors on the subsample spectra. They therefore apply no corrections for evolution of the galaxy population with redshift, and compute a single power spectrum for the entire sample.

2.2. Nonlinear Power Spectrum Models

The Q_{NL} model was introduced in Cole et al. (2005) as a fitting formula relating the linear and galaxy power spectra:

$$P_{gal}(k) = \frac{1 + Q_{NL}k^2}{1 + Ak} P_{lin}(k). \quad (4)$$

Both from examining mock catalogs and the HOD model analytically, Cole et al. (2005) fix $A = 1.4$ (redshift space) and $A = 1.7$ (real space) to model the ‘previrialization’ suppression of power on large scales (Lokas et al. 1996). Q_{NL} is considered a nuisance parameter whose amplitude is set both by nonlinearities in the matter field and by the scale-dependent bias of the galaxy sample.

Eisenstein et al. (2007b) examine in detail the damping of BAO features in the power spectrum through differential motion of pairs of tracers separated by the BAO scale. For our real space matter power spectrum we adopt their model

$$P_{smear}(k) = P_{lin}(k)e^{-k^2/2k_{BAO}^2} + P_{no\ wiggles}(k) \left(1 - e^{-k^2/2k_{BAO}^2}\right). \quad (5)$$

$P_{no\ wiggles}(k)$, defined by Eqn. 29 of Eisenstein & Hu (1998), is a smooth version of $P_{lin}(k)$ with the baryon oscillations removed. This means the baryon oscillations remaining in $P_{smear}(k)$ are damped completely for $k \gg k_{BAO}$ and unaltered for $k \ll k_{BAO}$. In § 6.1 we find $k_{BAO} = 0.14\ h\ \text{Mpc}^{-1}$ for our real space matter power spectrum, in agreement with the value reported in §6 of Eisenstein et al. (2007b); they also find that Eqn. 5 is a good approximation in redshift space, but with smaller k_{BAO} than in real space.

Percival et al. (2007) assume that both $P(k)$ and galaxy bias are linear, and show that the resulting fits to Ω_m for $0.01 < k < 0.06\ h\ \text{Mpc}^{-1}$ and $0.01 < k < 0.15\ h\ \text{Mpc}^{-1}$ are discrepant at the 2-3 σ level, demonstrating the need for nonlinear modeling to extract robust cosmological information from the observed galaxy power spectrum. The LRG power spectrum measured in Tegmark et al. (2006) requires a large nonlinear correction (see our Fig. 1). They adopt the model

$$P_{LRG}(k) = b^2 P_{smear}(k, k_{BAO} = 0.1) \frac{1 + Q_{NL}k^2}{1 + Ak} \quad (6)$$

with A fixed at 1.4 (M. Tegmark, private communication). Their best fit value for the nuisance parameter is $Q_{NL} = 30$. This model is only tested on power spectra of mock LRGs in *real* space and with no satellite LRGs. We will compare this model to our more realistic mock catalogs in redshift space.

2.3. Simplest HOD Model for the LRG Power Spectrum

As we describe in more detail in § 4.1, the HOD model specifies $P(N_{LRG}|M)$, the probability that N_{LRG} LRGs occupy a halo of mass M . Following Cooray & Sheth (2002), the nonlinear power spectrum can be separated into the contribution from pairs of galaxies occupying the same dark matter halo, $P^{1h}(k)$, and pairs occupying distinct dark matter halos, $P^{2h}(k)$. In the large scale limit of Eqn. 128 in Cooray & Sheth (2002), where the effect of finite size of halos can be safely neglected and the bias with respect to the dark matter can be approximated as linear, the LRG power spectrum is approximately

$$P_{LRG}(k) = P_{LRG}^{1h}(k) + P_{LRG}^{2h}(k) \quad (7)$$

$$P_{LRG}^{1h} = \int dM n(M) \frac{\langle N_{LRG}(N_{LRG} - 1) | M \rangle}{\bar{n}_{LRG}^2} \quad (8)$$

$$P_{LRG}^{2h}(k) = b_{LRG}^2 P_{DM}(k) \quad (9)$$

where $n(M)$ is the halo mass function, b_{LRG} is the large scale bias between the galaxies and matter, and $P_{DM}(k)$ is the nonlinear dark matter power spectrum. This simple model assumes that the effects of satellite galaxies are simply to reweight halo pairs traced by central galaxies in P^{2h} , and to add a simple shot noise term given by Eqn. 8. As we show below, this division of pairs works well in real space, and $P_{LRG}(k) - b_{LRG}^2 P_{DM}(k)$ is well described by P^{1h} . However, pairs of galaxies in redshift space cannot be cleanly described by this separation. For our HOD parameters, the halo velocity dispersion causes satellite galaxies to shift by $\langle \sigma^2 \rangle^{1/2} \sim 9 h^{-1}$ Mpc (with large tails) along the LOS. This shifting causes power to be shuffled between scales and even the largest scale modes along the LOS to be damped by the FOG smearing; Tinker (2007) provide a detailed treatment. Our approach will be to circumvent many of the difficulties in redshift space by eliminating the FOGs due to satellite galaxies from the density field before computing $P_{LRG}(k)$.

2.4. Modeling the Covariance Matrix

The covariance matrix for a set of band powers $P(k_{i=1..N})$ is defined by

$$C_{ij} = \langle (P(k_i) - \bar{P}(k_i))(P(k_j) - \bar{P}(k_j)) \rangle. \quad (10)$$

Hamilton et al. (2006) showed that the largest nonlinear contribution to the covariance matrix is a beat-coupling term proportional to the power on the largest scale of the survey. In any real survey of finite volume, the observed Fourier amplitudes are convolved with the

survey window function W_s :

$$\delta_{obs}(\mathbf{k}) = \int \delta(\mathbf{k}') W_s(\mathbf{k} - \mathbf{k}') \frac{d\mathbf{k}'}{(2\pi)^3}. \quad (11)$$

The beat coupling contribution arises because neighboring Fourier modes $\delta(\mathbf{k})\delta(-\mathbf{k} - \epsilon)$ are coupled by nonlinear growth to the beat mode $\delta(\epsilon)$. When the DC mode of the survey is positive, all modes are amplified; when it is negative, all modes are suppressed. This term can be large since the linear power spectrum drops so sharply with k . Hamilton et al. (2006) emphasize that this term does not contribute to the covariance of power measured from ensembles of traditional periodic box simulations, where the band power k_i is averaged over N_i complex modes by

$$\bar{P}(k_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta(\mathbf{k}_{ij}) \delta^*(\mathbf{k}_{ij}). \quad (12)$$

Here, the beat mode is $\mathbf{k}_{ij} - \mathbf{k}_{ij} = 0$, the DC mode. However, since our simulations allow the DC mode to vary (Sirko 2005), we will capture this term. We therefore model our covariance matrix as the sum of the usual diagonal Gaussian and shot noise contributions and the beat-coupling contribution:

$$C_{ij,model} = \frac{1}{N_i} \left(P(k_i) + \frac{\alpha_{shot}}{\bar{n}} \right)^2 \delta_{ij}^K + 4\beta_{beat} R_\alpha P(k_i) P(k_j) \delta_{DC}^2(z). \quad (13)$$

Here δ_{ij}^K is the Kronecker delta, and $\delta_{DC}^2(z) = P_L(k=0, z=0)/V_{sim} \times D^2(z)/D^2(z=0)$ is the variance of the DC mode linearly evolved to redshift z in the simulation volume $V_{sim} = L^3$ (Sirko 2005). In perturbation theory $R_\alpha \approx 2.62$ (Hamilton et al. 2006). We introduce α_{shot} and β_{beat} to allow for excess shot noise and variation in the amplitude of the beat coupling term, though we expect both parameters to be ~ 1 . We show in § 6 that Eqn. 13 provides a good model for both the dark matter and LRG covariance matrices. For the dark matter, the shot noise contribution is negligible.

Neyrinck et al. (2006) showed that Poisson fluctuations about the mean halo mass function introduce variance in the amplitude of the one-halo contribution to the dark matter power spectrum that can dominate the covariance matrix in the nonlinear regime. Since our reconstruction of the halo density field seeks to eliminate the one-halo contribution from galaxies, we expect the covariance to be smaller for the reconstructed halo density field than for the original galaxy sample.

2.5. Error Estimates on $P(k_i)$ and C_{ij}

For the dark matter, errors on the bandpowers are straightforward. They are simply the diagonal terms of the inverse covariance matrix divided by the number of simulations. We use the model covariance matrix in Eqn. 13 to compute the inverse:

$$\sigma_{P(k_i)}^2 = \frac{1}{N_{sim}} (\mathbf{C}_{model}^{-1})_{ii}. \quad (14)$$

Of course, the beat coupling leads to off-diagonal terms in both the covariance matrix and its inverse; these terms must be included when estimating errors on model parameters.

We also estimate an error on our estimate of the covariance matrix. We ignore correlations between the C_{ij} 's in our error estimates, though they are certainly present.

$$\sigma_{C_{ij}}^2 = \frac{1}{N_{sim}} \langle ((P(k_i) - \bar{P}(k_i))(P(k_j) - \bar{P}(k_j)) - C_{ij})^2 \rangle. \quad (15)$$

The situation for the mock catalogs is more tricky. For each of the $s = 1, \dots, N_{sim}$ N -body simulations we produce $m = 1, \dots, N_{mocks}$ mock catalogs using our fixed HOD parameters to reduce the shot noise contribution. We define

$$C_{ij,HOD} = \langle (P_{sm}(k_i) - \bar{P}_s(k_i))(P_{sm}(k_j) - \bar{P}_s(k_j)) \rangle \quad (16)$$

$$C_{ij,red} = \langle (\bar{P}_s(k_i) - \bar{P}(k_i))(\bar{P}_s(k_j) - \bar{P}(k_j)) \rangle \quad (17)$$

$$C_{ij,tot} = \langle (P_{sm}(k_i) - \bar{P}(k_i))(P_{sm}(k_j) - \bar{P}(k_j)) \rangle \quad (18)$$

$$C_{ij,tot} = C_{ij,HOD} + C_{ij,red}. \quad (19)$$

Here $P_{sm}(k_i)$ denotes the band power for mock catalog m populating simulation s , $\bar{P}_s(k_i)$ denotes the band power in a single simulation s averaged over the m mock catalogs populating s , and $\bar{P}(k_i)$ denotes a band power averaged over the entire set of $N_{mocks} \times N_{sim}$ catalogs. $C_{ij,tot}$ is the covariance matrix for this set of mock catalogs. $C_{ij,HOD}$ is the variance introduced by sampling the same matter density field with different mock catalog realizations and $C_{ij,red}$ is the reduced covariance of the power spectra from each simulation after averaging over N_{mocks} catalogs in each simulation. The expected error on bandpower $P(k_i)$ is then

$$P(k_i) = \langle P_{sm}(k_i) \rangle \quad (20)$$

$$\sigma_{P(k_i)}^2 = \frac{1}{N_{sim}} (\mathbf{C}_{model,red}^{-1})_{ii}. \quad (21)$$

We again neglect the covariance of the covariance matrix elements to estimate the errors on $C_{ij,tot}$, $C_{ij,red}$, and $C_{ij,HOD}$ from their variance over the simulations. These errors are only

used for fitting the parameters α_{shot} and β_{beat} , so inaccuracies in the errors should make our estimates noisier, but not biased.

3. Simulations

We use the publicly-available Tree-Particle-Mesh (TPM) code (Bode & Ostriker 2003) to run 42 periodic box simulations with the parameters selected in Chapter 3 of Reid (2008): $L_{box} = 558 h^{-1}$ Mpc and $N_p = 512^3$, corresponding to a particle mass $M_p = 1.43 \times 10^{11} M_\odot$. This resolution ensures that there are at least 50 particles per halo populated with an LRG for our best fit HOD parameters. We allow the slight change in the halo mass function in our lowest mass bin to be absorbed by a slight change in the best fit HOD parameters for central galaxies. As outlined in Appendix A, we performed a battery of tests to ensure that the large and small scale clustering and velocity statistics were unaffected by our choice of mass resolution for the simulation.

Initial conditions for the simulations were generated with the publicly available *ic* code (Sirko 2005). We updated the code to use the mt19937 “Mersenne Twister” random number generator (Matsumoto & Nishimura 1998) to select the initial density mode realizations. *ic* was designed to generate initial conditions for an ensemble of periodic box simulations that will match real-space statistical properties such as the mass variance in spheres and $\xi(r)$. For a periodic box simulation with comoving side length L_{uni} at $z = \infty$, the initial modes are drawn from a convolved power spectrum $P_{L_{uni}}(k)$ for which $P_{L_{uni}}(0)/L_{uni}^3$ is the variance of the DC mode in a volume equal to the initial periodic box volume. The DC mode is assumed to evolve linearly, and so its effects can be mimicked to first order in Lagrangian perturbation theory by a slight change in the cosmological parameters of the box as a function of the value of the DC mode in the realization (see Eqns. 19 - 22 of Sirko (2005)). The relation between the scale factor in the box cosmology, a_{box} , and the scale factor in the universe cosmology, a_{uni} , is

$$\frac{a_{box}}{a_{uni}} = 1 - \frac{1}{3} \frac{D_{uni}(a_{uni})}{D_{uni}(1)} \Delta_o \quad (22)$$

where Δ_o is the amplitude of the DC mode at $a_{uni} = 1$, and the growth function D_{uni} is evaluated using the universe cosmology. In this scenario, each simulation in the ensemble represents an equal initial comoving volume, L_{uni}^3 . Once the universe has evolved to scale factor a_{uni} , the comoving size of the box is $L_{box} = L_{uni}(a_{box}/a_{uni})$. Because overdense regions expand more slowly than underdense regions, we must weight the simulations by $(a_{box}/a_{uni})^3$ to obtain volume-averaged quantities. This improvement over a set of simulations for which

the DC mode is precisely 0 is crucial to the analysis in this paper, since we wish to extract accurate matter and halo power spectra as the density field enters the nonlinear regime, along with accurate estimates of band power covariances.

We output the dark matter particle positions and velocities at time steps nearest to universe redshifts $z_{NEAR} = 0.235$, $z_{MID} = 0.342$, and $z_{FAR} = 0.421$, the galaxy-weighted mean redshifts of the three redshift subsamples analyzed in Tegmark et al. (2006). Eqn. 22 shows that for a snapshot of the universe at fixed redshift, each simulation should be output at a different z_{box} that increases with Δ_o . Since the TPM code outputs particle data at the nearest whole time step to the desired redshift, the difference between the target and output redshift will vary slightly with Δ_o . We correct for this effect exactly in the linear regime by scaling the power spectra from each simulation by $(D(z_{target})/D(z_{output}))^2$ before computing averaged quantities. Without this scaling, we get the same average dark matter power spectra; changes in the covariance estimates are well below 1%. Any nonlinear corrections to this scaling will have an even smaller effect, and so we safely neglect them.

3.1. Calculating Power Spectra

Because of the low number density of LRGs, the shot noise correction is large compared to the $P(k)$ of the continuous density field sampled by the LRGs. We use an FFT with $N = 512^3$ points, and present the power spectrum out to $k = 0.4 h \text{ Mpc}^{-1} = 0.14k_N$, where $k_N = \pi/\Delta_{grid}$ is the Nyquist frequency. Comparison with an $N = 1024^3$ grid showed agreement at the $\sim 10^{-6}$ level. Following the work of Hockney & Eastwood (1981), Jing (2005), and Baugh & Efstathiou (1994), mock LRGs are distributed on the FFT grid using the triangular-shaped cloud (TSC), and the power spectrum is estimated at each \mathbf{k} by (Jing 2005)

$$P(\mathbf{k}) = \frac{|\delta_{LRG,FFT}(\mathbf{k})|^2 - P_{shot,TSC}(\mathbf{k})}{W_{TSC}^2(\mathbf{k})} \quad (23)$$

$$P_{shot,TSC}(\mathbf{k}) = \frac{1}{\bar{n}} \Pi_i \left[1 - \sin^2 \left(\frac{\pi k_i}{2k_N} \right) + \frac{2}{15} \sin^4 \left(\frac{\pi k_i}{2k_N} \right) \right] \quad (24)$$

$$W_{TSC}(\mathbf{k}) = \Pi_i \left[\text{sinc} \left(\frac{\pi k_i}{2k_N} \right) \right]^3, \quad (25)$$

where $\delta_{LRG,FFT}(\mathbf{k})$ is the FFT of the LRG overdensity field and k_i are the Cartesian components of \mathbf{k} . Eqn. 23 corrects for FFT aliasing in the limit that $P(\mathbf{k} + \mathbf{k}_N)$ is shot noise dominated, which is the case here. Finally, we average the values of Eqn. 23 over k-bands

with $\Delta k_{box} = 0.0113$.

3.2. Halo Catalogs

We use the spherical overdensity (SO) halo finder code described in Tinker et al. (2008) with $\Delta = 200\rho_b$ to generate halo catalogs. Fig. 2 shows the ratio of the volume-weighted SO mass function in our simulation set to the fitting function of Tinker et al. (2008) down to $M = 7.15 \times 10^{12} M_\odot$. There is good agreement at the level of accuracy claimed for their analytic fits ($\sim 5\%$). The overabundance of halos in the lowest mass bins is likely due to the small number of particles per halo in those bins. This slight modification of the mass function is unimportant for our purposes, since there are no satellite LRGs in these mass bins, and changes in the mass function are degenerate with changes to the probability of hosting a central galaxy as a function of halo mass, $N_{cen}(M)$.

4. Mock Catalogs

We use the technique described in Reid & Spergel (2008) to produce mock catalogs for the NEAR ($0.155 < z < 0.300$), MID ($0.300 < z < 0.380$), and FAR ($0.380 < z < 0.474$) redshift subsamples of Tegmark et al. (2006). These samples contain roughly equal numbers of galaxies, and their power spectra have roughly the same amplitude on large scales. We first review our HOD modeling assumptions, described in more detail in Reid & Spergel (2008).

4.1. HOD Model

The Halo Occupation Distribution (HOD) model assumes that the probability $P(N_{LRG}|M)$ of N_{LRG} LRGs occupying a dark matter halo of mass M at redshift z depends only on the halo mass (for a review, see Cooray & Sheth 2002). However, the application of the HOD formalism requires that we make several further assumptions about $P(N_{LRG}|M)$. Detailed studies of dark matter halos and subhalos suggest a division of galaxies into central and satellite galaxies (Kravtsov et al. 2004). The central galaxies are assumed to sit at the halo center, consistent with the observation that most ($\sim 80\%$) of the brightest cluster LRGs are

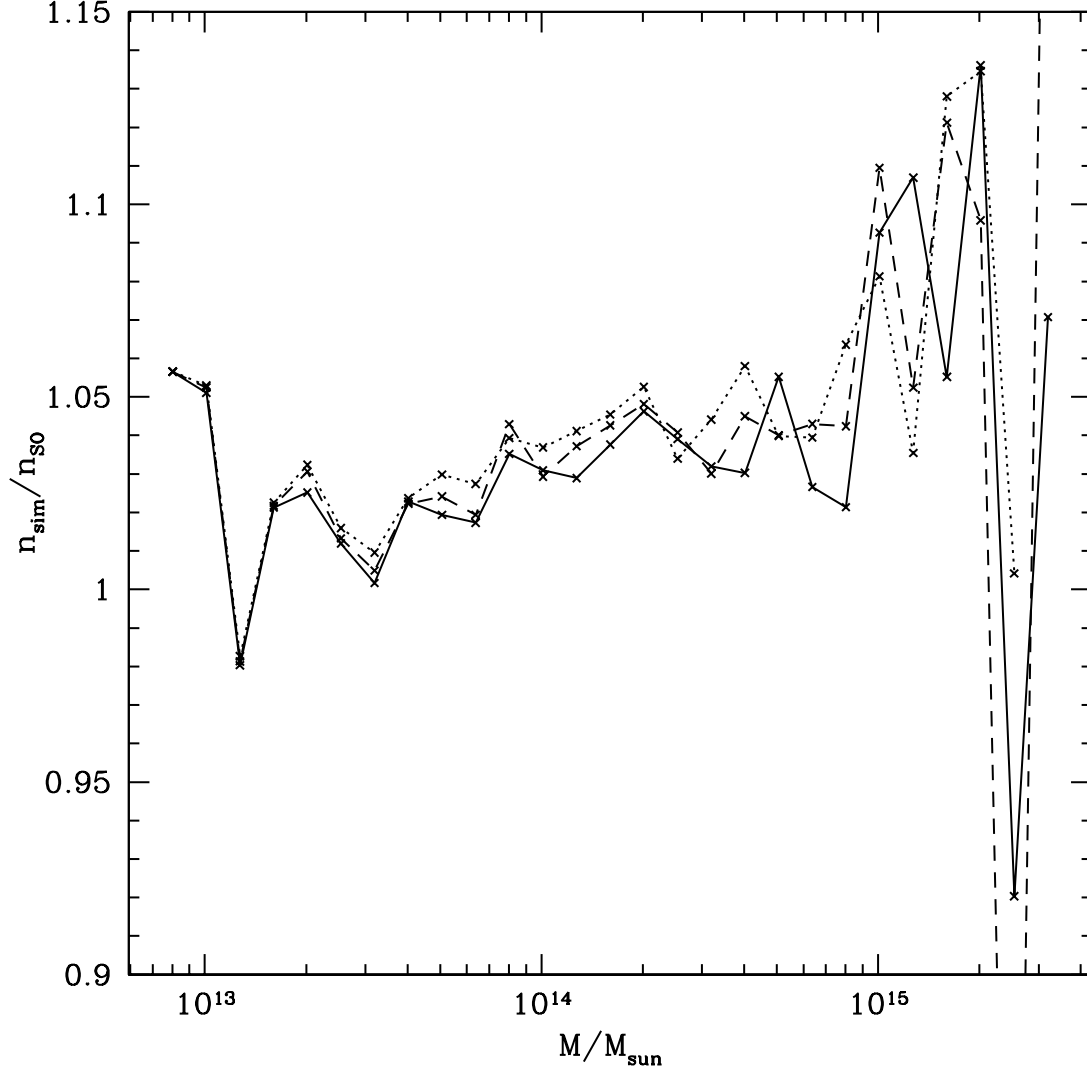


Fig. 2.— The ratio of the volume-weighted SO mass function from our simulations with the spline fit given in Tinker et al. (2008) for the NEAR (solid line), MID (dashed line), and FAR (dotted line) in bins of $\Delta \log_{10} M = 0.1$. The agreement is within the stated accuracy of the Tinker et al. (2008) spline fit ($\sim 5\%$).

found within $0.2r_{vir}$ of the center of the cluster potential well as traced by X-rays. Satellite galaxies occur in the more massive halos already containing a central galaxy. In high resolution simulations they can be directly associated with dark matter subhalos (Vale & Ostriker 2006); here we will assume they have the same distribution as the halo dark matter. We will use these functional forms with the five free parameters M_{min} , σ_{logM} , M_1 , M_{cut} , and α to describe the mass dependence of the average halo occupation as a function of halo mass M :

$$\langle N(M) \rangle = \langle N_{cen} \rangle (1 + \langle N_{sat} \rangle) \quad (26)$$

$$\langle N_{cen} \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log_{10} M - \log_{10} M_{min}}{\sigma_{logM}} \right) \right] \quad (27)$$

$$\langle N_{sat} \rangle = \left(\frac{M - M_{cut}}{M_1} \right)^\alpha. \quad (28)$$

Halos are populated with a central galaxy with probability $N_{cen}(M)$. Central galaxies are placed at the center of their host halos and assigned the peculiar velocity of their halos. Halos with a central galaxy are populated with N_{sat} galaxies, where $P(N_{sat}|N(M))$ is drawn from a Poisson distribution. The position and velocity of a satellite galaxy is taken to be that of a randomly selected dark matter particle halo member. We assign comoving redshift space position s to an object in our mock catalogs using the conversion at z_{box} , the redshift at which the simulation data were output:

$$s = x_{LOS} + (1 + z_{box})v_p/H(z_{box}) \quad (29)$$

where x_{LOS} is the comoving distance along the line of sight in real space.

4.2. Measurements of N_{CiC} from SDSS and HOD parameter results

The HOD model specified in § 4.1 determines the number of groups with $n_{sat} = 0, 1, 2, \dots$ satellite galaxies. In Reid & Spergel (2008) we present the measurement of the SDSS Counts-in-Cylinders (CiC) group multiplicity function $N_{CiC}(n_{sat})$, calibrate its relation to the true group multiplicity function, and use a maximum likelihood analysis to derive the parameters of $N_{sat}(M)$ in Eqn. 28. The parameters of $N_{cen}(M)$ in Eqn. 27 are constrained by the observed LRG number density and large scale clustering amplitude. To derive HOD parameters for the mock catalogs used in this paper, we follow the technique of Reid & Spergel (2008) exactly, but we must first measure $N_{CiC}(n_{sat})$ separately for our three redshift subsamples.

In Table 1 we report our measurement of $N_{CiC}(n_{sat})$ from the SDSS DR4+ LRG sample for the NEAR, MID, and FAR redshift subsamples. We introduce a few minor changes in

this measurement from Reid & Spergel (2008). First, we include a boundary set of galaxies in the redshift direction, since FOGs may widely separate nearby pairs of LRGs; we verified that this has a small effect on the resulting CiC group multiplicity. This buffer is $\Delta z = 0.007$ at the low redshift end of the sample and $\Delta z = 0.009$ at the high redshift; that is, equal to the maximum redshift separation for a CiC pair of galaxies at the sample boundary. This causes the redshift ranges of the subsamples for which we measure $N_{CiC}(n_{sat})$ to be slightly different than the Tegmark et al. (2006) NEAR, MID, and FAR samples. Because the color cuts used to select the LRG sample produces a complex radial selection function, the effective number density of the two samples is slightly different. As discussed in Reid & Spergel (2008), the main difference between our sample number densities (Column 4 of Table 2) and the result of integrating the Zehavi et al. (2005) model for the redshift dependence of $\bar{n}_{LRG}(z)$ (Column 5 of Table 2) comes from our careful inclusion of objects from the imaging sample. In the MID sample, this is an 8% increase in the number density. We have verified that once the imaging galaxies are accounted for along with the difference in redshift range of the NEAR and FAR samples, our simple number density estimate $n_{LRG} = N_{sample}/V_{sample}$ is in agreement with the expectations from the Zehavi et al. (2005) model. Second, our redshift indicator used for objects without spectra must be altered, as the 4000 Å break moves from the g to the r band at $z \simeq 0.4$ (see Fig. 4 of Eisenstein et al. 2001). For our FAR sample we use the $r - i$ color as a redshift indicator. We use the observed $N_{CiC}(n_{sat})$ for the NEAR, MID, and FAR subsamples to derive the subsample satellite HOD parameters M_{cut} , M_1 , and α in Table 2 using the maximum likelihood technique presented in Reid & Spergel (2008). Since Tegmark et al. (2006) find that the NEAR, MID, and FAR subsamples have consistent $P(k)$'s, we vary σ_{logM} to match each redshift subsample to the large scale amplitude of the combined $P(k)$ reported in Tegmark et al. (2006). M_{min} is determined by \bar{n}_{sim} . Fig. 3 shows the agreement of the large scale $P(k)$ with Tegmark et al. (2006) when we apply their FOG compression algorithm to our mock catalogs.

5. Reconstructing the Halo Density field

In the CiC technique detailed in Reid & Spergel (2008), two galaxies are considered neighbors when their transverse comoving separation satisfies $\Delta r_{\perp} \leq 0.8 h^{-1}$ Mpc and their redshifts satisfy $\Delta z/(1+z) \leq \Delta v_p/c = 0.006$. A cylinder should be a good approximation to the density contours of satellites surrounding central galaxies in redshift space, as long as the satellite velocity is uncorrelated with its distance from the halo center and the relative velocity dominates the separation of central and satellite objects along the line of sight. Galaxies are then grouped with their neighbors by a FoF algorithm. The reconstructed halo

Table 1. The final $N_{CiC}(n_{sat})$ group multiplicity function following the method in Reid & Spergel (2008) for our NEAR, MID, and FAR LRG subsamples.

n_{sat}	$N_{CiC,NEAR}(n)$	$N_{CiC,MID}(n)$	$N_{CiC,FAR}(n)$
0	22921.71	24537.81	19109.71
1	1372.63	1301.29	664.28
2	170.01	153.40	61.94
3	41.85	25.59	6.08
4	15.16	9.07	2.04
5	2.11	2.04	1.00
6	1.01	1.01	0.00
7	1.02	0.00	0.00
8	0.03	0.00	0.00

Table 2. Mock catalog parameters. Masses in units of $10^{14} M_{\odot}$ are for SO halos with $\Delta = 200\rho_b$; number densities are in units of $10^{-4} (h^{-1} \text{ Mpc})^{-3}$. $z_{SDSS,CiC}$ is the redshift range of the SDSS LRG subsample and \bar{z}_{sim} is the average simulation z_{uni} . \bar{n}_{sim} is the mock catalog number density while \bar{n}_{model} is the expected number density obtained by integrating the Zehavi et al. (2005) model over the $z_{SDSS,CiC}$ range. P^{1h} is computed from Eqn. 8 in units of $(h^{-1} \text{ Mpc})^3$.

Sample	$z_{SDSS,CiC}$	\bar{z}_{sim}	\bar{n}_{sim}	\bar{n}_{model}	σ_{logM}	M_{min}	M_{cut}	M_1	α	P^{1h}
NEAR	0.162 - 0.300	0.2347	1.05	0.982	0.6	0.78	0.49	5.87	1.16	1759
MID	0.300 - 0.380	0.3420	0.962	0.890	0.6	0.76	0.57	6.29	1.05	1480
FAR	0.380 - 0.465	0.4216	0.470	0.422	0.9	2.27	1.38	5.97	0.78	2312

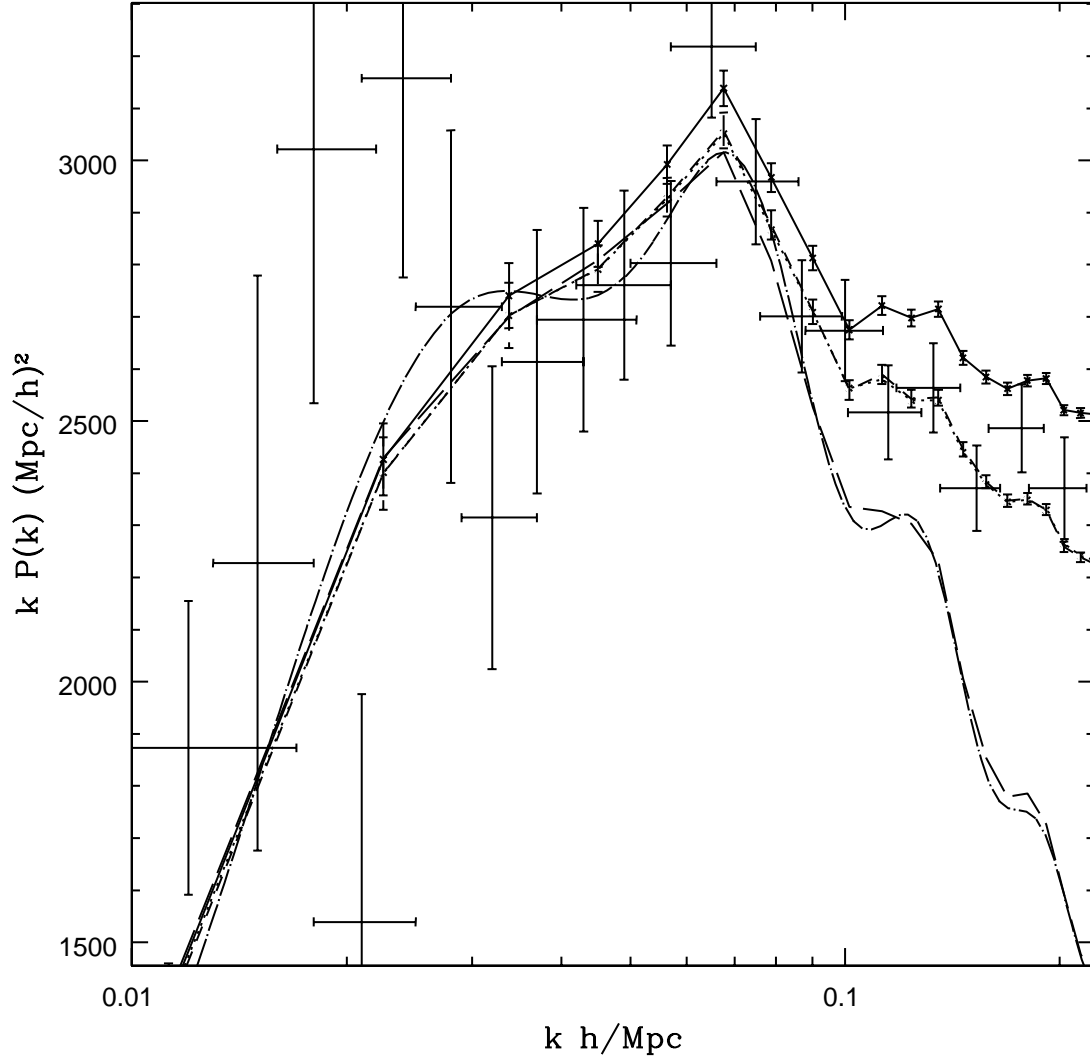


Fig. 3.— $kP(k)$ for the NEAR (dotted curve), MID (dashed curve), and FAR (solid curve) mock galaxy samples with the FOG compression algorithm of Tegmark et al. (2006) applied compared with their observed power spectrum (points with large error bars). Our NEAR and MID samples are indistinguishable, so we show error bars only for the MID and FAR samples. Our error bars are derived from the diagonal elements of the inverse $C_{ij,red}$ matrix. We scale the real space $P(k)$ of Tegmark et al. (2006) by $(0.8)^{-1}$ to approximate the redshift space monopole (i.e., angle averaged) power spectrum; this is the relation between the real and redshift monopole spectra if one neglects the small contributions from the quadrupole and hexadecapole detailed in their §A3. For comparison, the long dashed curve shows our simulation initial conditions drawn from the convolved power spectrum according to the *ic* algorithm, and the smooth dot-dashed curve shows the linear power spectrum for the cosmological parameters adopted in this study.

density field is defined by the superposition of the centers of mass of the CiC groups. The CiC parameters were established as a balance between completeness and contamination in identifying pairs of galaxies at an early stage of this work based on FoF halo catalogs. While the parameters and details of the method described here are sufficient for approximately recovering the halo density field power spectrum, the method could almost certainly be improved to more accurately recover groups of galaxies residing in the same dark matter halo.

6. Results

6.1. Dark Matter

6.1.1. Matter Power Spectrum

Fig. 4 shows the ratio of the matter power spectrum to the power spectrum of the simulation initial conditions scaled by the expected linear growth $D^2(z)$ for our three redshift samples; the results at different redshifts are highly covariant since they are measured from the same set of simulations at relatively small separations in time. Error bars are computed from the inverse of the model covariance matrix in Eqn. 13 with $\beta_{beat} = 1$ fixed.

As expected, the nonlinear correction grows as the redshift decreases. The nonlinear evolution generates power that smoothly increases with k , and also damps the baryon oscillations. The halofit nonlinear correction (Smith et al. 2003) (dotted curve, evaluated at z_{MID}) underestimates both the baryon wobble suppression and the amplitude of the smooth increase in power; Crocce & Scoccimarro (2008) also find a disagreement between their N -body simulation results and the halofit nonlinear correction.

Our fitting function to the nonlinear matter power spectrum $P_{DM}(k)$ aims to capture both the smearing of the acoustic peaks and a smooth increase in power with k . We follow Eisenstein et al. (2007b) in defining $P_{smear}(k)$ (Eqn. 5), but adopt a slightly different fitting function for the smooth correction. Our model is

$$P_{DM}(k) = P_{smear}(k; k_{BAO}) (a_0 + a_1 k + a_2 k^2 + a_3 k^3). \quad (30)$$

For the MID sample with 35 bandpowers and 5 parameters, $\chi^2 = 21$. We find $k_{BAO} = 0.14$, in good agreement with the values reported in Eisenstein et al. (2007b). In Fig. 5 we show that the $P_{smear}(k)$ term accounts for the baryonic features in $P_{DM}(k)$, and the polynomial

in k adequately fits the smooth correction for the MID sample, while halofit underestimates the smooth correction by $\sim 4\%$ at $k = 0.2$. The NEAR and FAR fits are similar. In Table 3 we list fits for the NEAR, MID, and FAR power spectra out to a maximum k of 0.2 and $0.4 \ h \text{ Mpc}^{-1}$. When $k_{max,fit} = 0.2 \ h \text{ Mpc}^{-1}$, only the first three terms in the polynomial expansion are necessary for a good fit.

6.1.2. Covariance Matrix

Fig. 6 shows diagonal elements of the normalized covariance matrix, $C_{ii}/\bar{P}(k_i)^2$. We estimate the errors from the diagonal variances (Eqn. 15); these may not capture the true errors since off-diagonal elements should be present in the 8-point function as well. Nevertheless, when we use these error estimates to compute χ^2 for the model in Eqn. 13, we find $\chi^2 = 1600$ for 1296 degrees of freedom ($0 \leq k \leq 0.4$); if we restrict the covariance matrix to the 196 elements with both k bands between 0.056 and 0.21, we find $\chi^2 = 262$. In this case there are no free parameters and we deem the model a reasonable fit. If we allow the amplitude of the beat coupling term to vary, we find a best fit value 0.96 for the full matrix ($\chi^2 = 1564$) and 0.84 for the $0.056 \leq k \leq 0.21$ subsample ($\chi^2 = 218$). We note that for the DC mode realizations of our 42 simulations, the variance is a factor of 0.83 lower than the expected variance (in agreement with the expected random variation for a single mode, $N_{sim}^{-1/2} = 15\%$). We conclude that Eqn. 13 is an excellent fit to our dark matter covariance matrix.

Another point of interest in Fig. 6 is that the inverse of the diagonal elements of the inverse covariance matrix are nearly equal to the Gaussian expectation (dashed curve). This means that while the beat-coupling does not introduce additional errors on the measurement of the bandpowers, the measured values will be covariant. However, for a beat-coupling term in the form of Eqn. 13, only information on the overall amplitude of $P(k)$ is lost. However, large scale structure analyses traditionally marginalize over the amplitude of $P(k)$, since σ_8 and bias are degenerate.

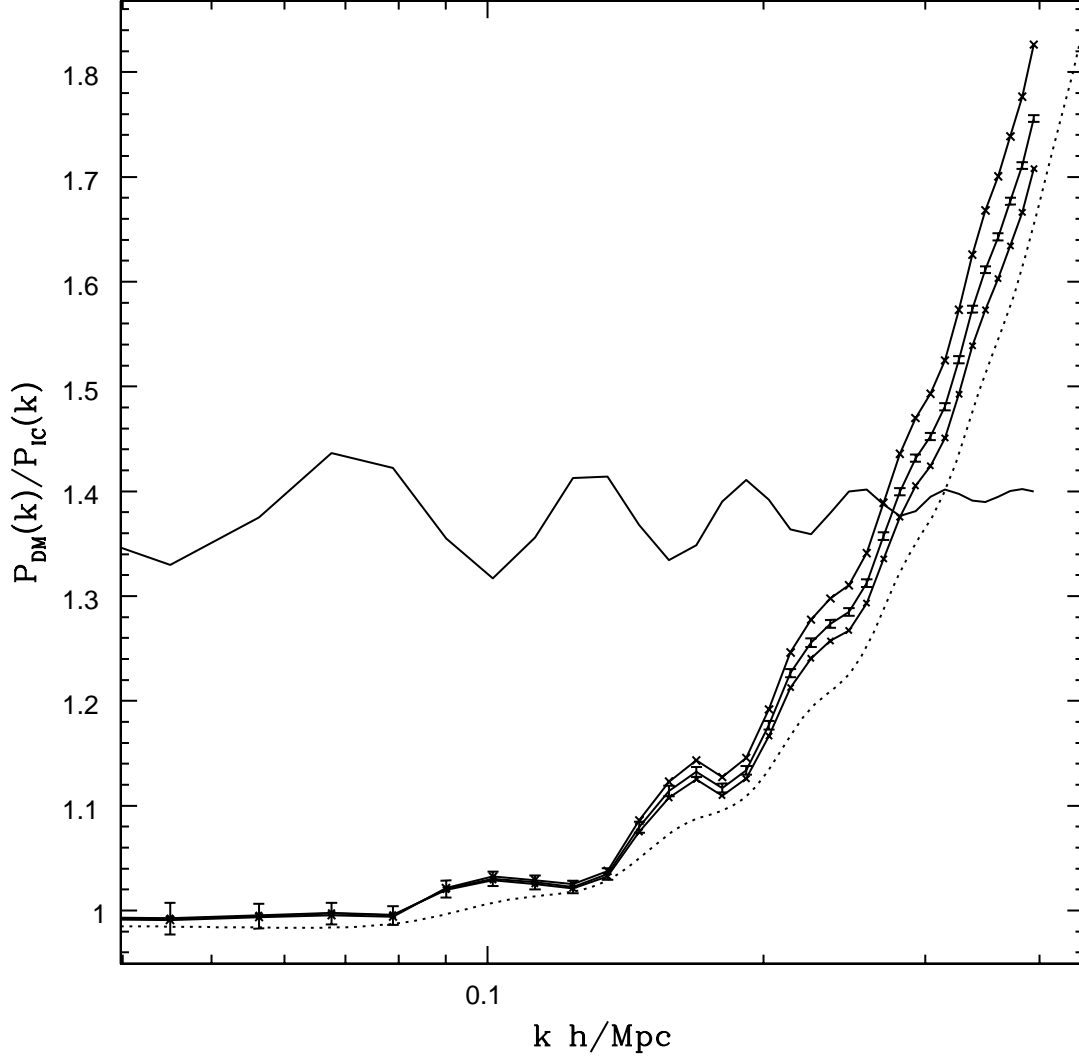


Fig. 4.— The ratio of the volume-averaged matter power spectra at $z_{FAR} = 0.421$, $z_{MID} = 0.342$, and $z_{NEAR} = 0.235$ to the input power spectrum scaled by $D^2(z)/D^2(z=0)$; the amplitude of the nonlinear power spectrum increases as the redshift decreases. The nonlinear correction to the matter power spectrum measured from our simulations is larger than expected from halofit (Smith et al. 2003), shown as the dotted curve and evaluated at z_{MID} . We also overlay $P_{IC}(k)/P_{\text{no wiggles}}(k) + 0.4$ (solid curve oscillating about 1.4) to indicate the location of the baryon wiggles in the initial power spectrum. The nonlinear evolution generates power that smoothly increases with k and damps the baryon oscillations. Error bars are shown only for the MID sample for clarity and are estimated from the inverse of the model covariance matrix defined in Eqn. 13, where the shot noise is negligible and $\beta_{beat} = 1$; the beat-coupling term in Eqn. 13 correlates the bandpowers.

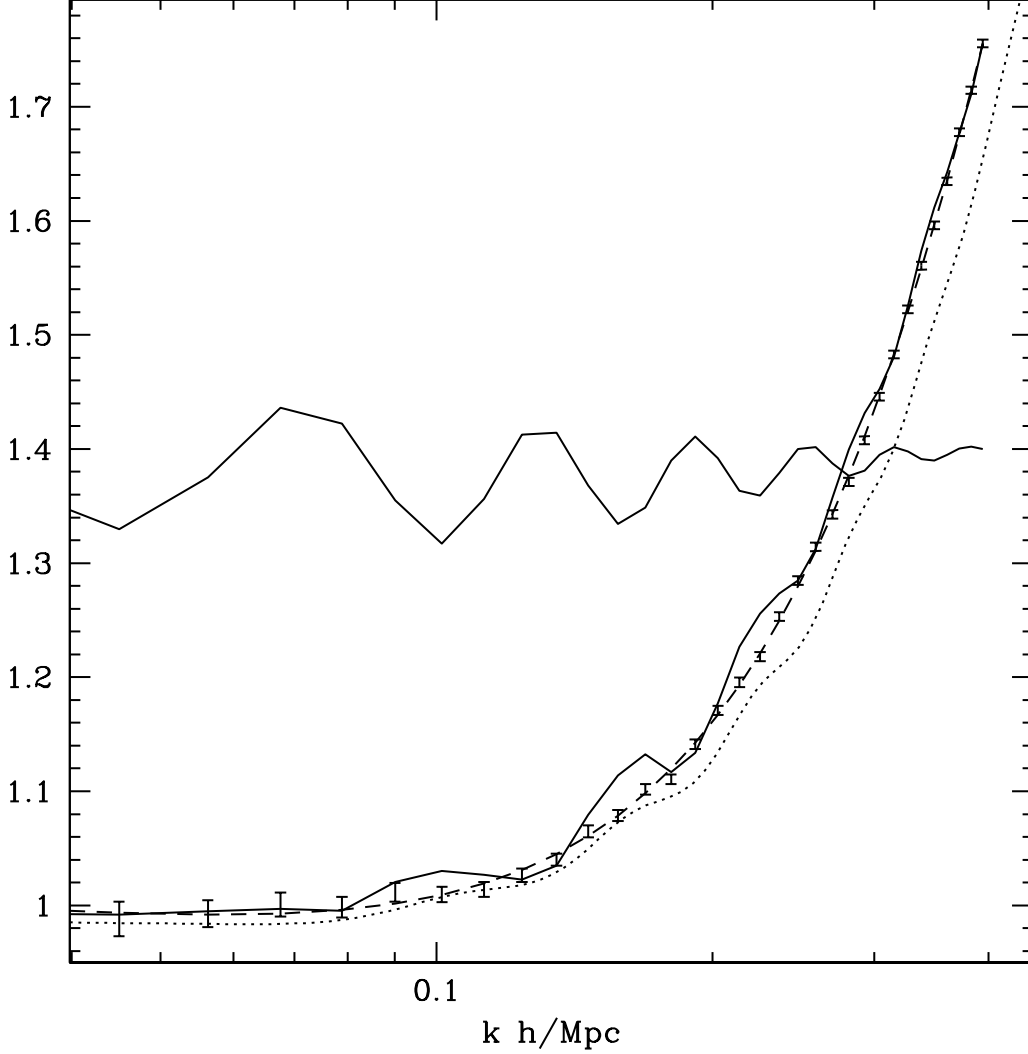


Fig. 5.— $P_{DM}(k)/P_{IC}(k)$ for the MID sample (solid curve, as in Fig. 4), $P_{DM}(k)/P_{\text{smeared}}(k)$ for $k_{BAO} = 0.14$ (points with error bars), and our polynomial fit $1.026 - 1.199k + 11.06k^2 - 8.426k^3$ (dashed curve). We also overlay $P_{IC}(k)/P_{\text{no wiggles}}(k) + 0.4$ (solid curve oscillating about 1.4) to indicate the location of the baryon wiggles in the initial power spectrum. Using error bars derived from the inverse of the model covariance matrix defined in Eqn. 13, $\chi^2 = 21$ for 35 bandpowers and 5 parameters. Comparison of the solid and dashed curves shows that $P_{\text{smeared}}(k)$ adequately models the BAO features, and that our third-order polynomial in k is sufficient for $k \leq 0.4 \, h \, \text{Mpc}^{-1}$. The halofit (Smith et al. 2003) correction (dotted curve) shows better agreement with the smooth portion of the nonlinear correction, though halofit underestimates the nonlinear power by 4% at $k = 0.2 \, h \, \text{Mpc}^{-1}$.

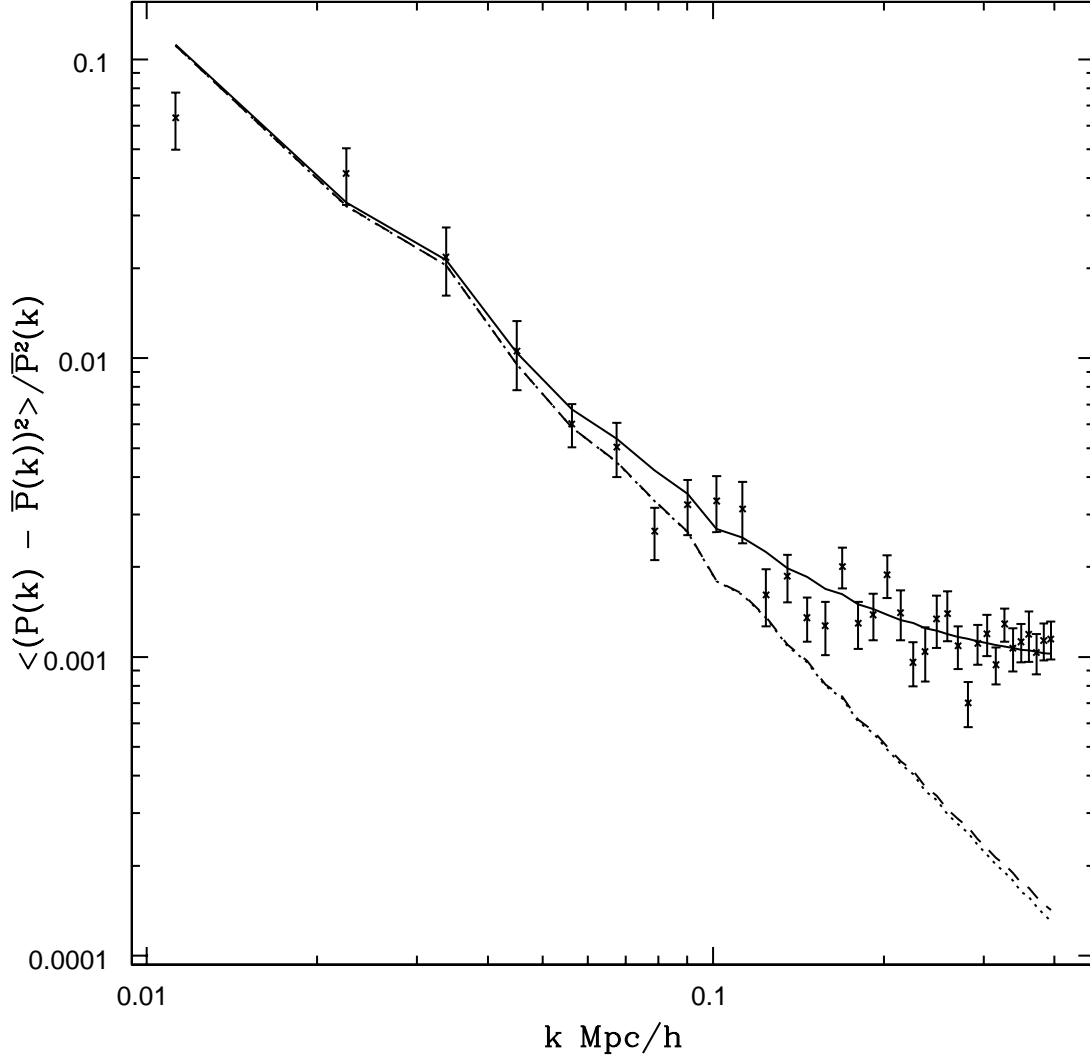


Fig. 6.— The points with error bars show the measured diagonal elements of the covariance matrix. Error bars are estimated from the variance of C_{ii} across our simulations. The dotted curve shows the Gaussian prediction for the diagonal covariance, the inverse of the number of complex modes in each linearly-spaced k -bin. The solid curve shows the sum of the Gaussian term and the beat-coupling term in Hamilton et al. (2006), which provides a good model for the full covariance matrix. The dashed curve shows the inverse of the diagonal elements of the inverse covariance matrix. We find that while the errors on the bandpowers are very close to the Gaussian expectation, the beat-coupling terms in the covariance matrix mean the bandpowers are correlated.

6.2. Mock Catalogs

6.2.1. Power Spectra

Fig. 3 shows the agreement between our NEAR, MID, and FAR power spectra and the measurements presented in Tegmark et al. (2006). We used this comparison to set by eye the large scale bias of our mock catalogs through $\sigma_{\log M}$, the width of $N_{cen}(M)$ in Eqn. 27. The small discrepancy between the normalizations of the FAR and other catalogs at small k could be eliminated with a slight variation in $\sigma_{\log M, FAR}$. However, the FAR sample clearly has a different shape, and is consistent with trends in Fig. 6 of Tegmark et al. (2006), though in the SDSS sample the significance of these trends is less clear. Our FAR sample has $\sim 3\%$ more power at $k = 0.09$ and $\sim 10\%$ more power at $k = 0.2$ than the NEAR and MID samples.

Having established the agreement between our mock catalogs and the Tegmark et al. (2006) observed $P_{LRG}(k)$, we now attempt to isolate the sources of nonlinearity in our catalogs and demonstrate that the power spectrum of the reconstructed halo density field is the best tracer of the underlying linear spectrum. In Fig. 7 we analyze in detail the MID sample, while Figs. 8 and 9 show that the NEAR and FAR sample behave similarly. The first major result is that there is no detectable deviation from a constant bias between the central LRGs and the dark matter in real space for $k \leq 0.1$, and the deviation at $k = 0.2$ is $\leq 2\%$ for the MID and FAR subsamples; the discrepancy for the NEAR sample is $\sim 1\%$ at $k = 0.1$ and 5% at $k = 0.2$ (solid curves with the lowest amplitude as $k \rightarrow 0$). The lower left panel of Fig. 7 demonstrates that the main effect of using redshift space coordinates for the central LRGs is the further damping of the BAO signatures. There is also a nonmonotonic but smooth variation that is $\leq 4\%$ between $k = 0$ and $k = 0.4$. In the $k \leq 0.2 h \text{ Mpc}^{-1}$ regime, the direction of the deviation is opposite of that between the real space central LRGs and underlying dark matter, so that the redshift space central LRG power spectrum is nearly linearly related to the real space matter power spectrum at $k \leq 0.2 h \text{ Mpc}^{-1}$.

The inclusion of satellite LRGs in real space has two effects. The linear bias is increased because the satellites upweight only the most massive, more highly biased halos traced by the LRGs. Secondly, the satellite galaxies add a shot noise given by Eqn. 8. The upper right panel of Fig. 7 shows that the difference between the $P(k)$ including satellites and $P(k)$ of the central LRGs only is well described by these two effects, with a relative bias $b_{rel} = 1.042$ and $P^{1h} \approx 1460 (h^{-1} \text{ Mpc})^3$. This latter value is in good agreement with $1480 (h^{-1} \text{ Mpc})^3$, the value computed directly from Eqn. 8 using our input HOD parameters. As $k \rightarrow 0.4$, the amplitude of the difference slightly diminishes due to the width of the satellite

LRG density profile within the halos. P^{1h} can be $\sim 7 - 11\%$ of $P_{LRG}(k)$ at $k = 0.1$, and $\sim 15 - 23\%$ at $k = 0.2$ (see Table 2). If uncorrected for, this will be the dominant source of nonlinearity for LRGs owing to their low number density, which appears in the denominator of Eqn. 8. Moreover, since the expected P^{1h} varies across the NEAR, MID, and FAR samples, the power spectrum shape will necessarily vary with redshift, as is evident in Fig. 3.

We now consider the implications of the FOG-compression algorithm of Tegmark et al. (2006), which was designed to recover the real space $P_{LRG}(k)$. Figs. 7 through 9 demonstrate that their compression algorithm (Eqn. 3) is overly aggressive and adds more one-halo power than is present in the real space mock catalog (dash-dot curves). This is the main source of the large nonlinear correction required in Tegmark et al. (2006). However, Fig. 10 shows that the nonlinear correction model of Tegmark et al. (2006) in Eqn. 6 accurately describes their resulting power spectrum for each of the redshift subsamples, so that the resulting cosmological constraints should not be biased. The best fit Λ CDM - Q_{NL} model in Tegmark et al. (2006) (long-dashed curve of Fig. 10) is consistent with being a weighted average over the NEAR, MID, and FAR subsample mock catalog power spectra. The dot-dashed curve in Fig. 10 shows the ratio of the linear $P(k)$ for their best fit cosmological parameters with $P_{IC}(k)$ from our simulations. These two models differ at a level larger than the statistical error bars shown in Fig. 1, even if we restrict the analysis to $k < 0.1$, and certainly if we can use the measurements between $k = 0.1$ and $k = 0.2$. However, the Q_{NL} parameter is degenerate with this difference in power spectra, so that with this nonlinear correction model we cannot distinguish the two linear spectra.

In redshift space the effect of the satellite galaxies on the angle-averaged power spectrum is more complicated. At high k the ratio $P(k)/P_{DM}(k)$ turns over as the suppression of power by the FOGs becomes more important than the addition of one-halo power. In the bottom right panel of Fig. 7 we see that the difference between the all LRG and central LRG power spectra is no longer well-approximated by the constant $P^{1h} \sim 1480 (h^{-1} \text{ Mpc})^3$ from Eqn. 8. The one-halo pair separation in redshift space is large ($\sim 9 h^{-1} \text{ Mpc}$ but with broad tails), so that the neat division between one-halo and two-halo pairs in real space no longer holds, and power will be transferred between k -bands in a complicated way. The reconstructed halo density field method removes the satellite galaxies before computing $P(k)$, which allows us to avoid the necessarily complicated FOG modeling (Tinker 2007).

The power spectrum of the reconstructed halo density field is very similar to the redshift space power spectrum of the central LRGs for the NEAR, MID, and FAR subsamples, as one

would expect if the reconstruction is sufficiently accurate. The slight difference is evident in the bottom left panel of Fig. 7, and may be because our method puts the reconstructed halo at the center of mass of the CiC group, which will leave residual FOG smearing. An improved method may be to put the reconstructed halo at the position of the brightest LRG, which should be closest to the halo center and have a smaller velocity with respect to its host halo. For $k \leq 0.1$ the reconstructed halo density field does not detectably deviate from the shape of the underlying matter power spectrum for the NEAR and MID samples; for the FAR sample, the deviation is $\sim 1\%$. In contrast, the FOG-compressed sample deviates from a constant bias at the 6%, 7%, and 10% level between $k = 0.05$ and $k = 0.1$. The reconstructed halo density field has only small deviations from a scale-independent bias out to $k = 0.2$: the deviation is 4%, 2.8%, and 2.5% for the NEAR, MID, and FAR sample. The systematics out to $k = 0.2$ should therefore be small enough for use in cosmological analyses. In contrast, the FOG-compressed spectra deviate from a constant bias at the 19%, 20% and 30% levels for the NEAR, MID, and FAR samples; this is a factor of ~ 5 larger than the statistical errors on the bandpowers in Tegmark et al. (2006) between $k = 0.1$ and $k = 0.2$. The nonlinear correction must therefore be extremely well-calibrated to extract any cosmological information in this regime.

We have shown that the nonlinear correction between the reconstructed halo density field and the matter power spectrum is very small. The small corrections we have found are well fit away from $k = 0$ by a polynomial. We list best fit polynomial parameters in Table 3 in Appendix B. The amplitude of this correction in our fiducial cosmology should provide a conservative estimate of the uncertainty in the correction as the cosmology is varied within the space allowed by the latest WMAP analysis Komatsu et al. (2008): $\sim 1\%$ for $k \leq 0.1$ and $\sim 4\%$ between $k = 0.1$ and $k = 0.2$. Finally, to recover the linear power spectrum, one must know the amplitude of the BAO suppression and the degree of nonlinearity in the matter power spectrum as a function of the cosmological parameters. Several groups are addressing this issue (e.g., Eisenstein et al. 2007b; Habib et al. 2007).

6.2.2. Covariance Matrices

We use the scatter in the power spectrum measurements in the simulations to estimate the covariance matrix for our mock catalogs. We fit the measurements of the four-point function to a physically motivated model:

$$B_{ij} = 4R_\alpha P(k_i)P(k_j)\delta_{DC}^2(z) \quad (31)$$

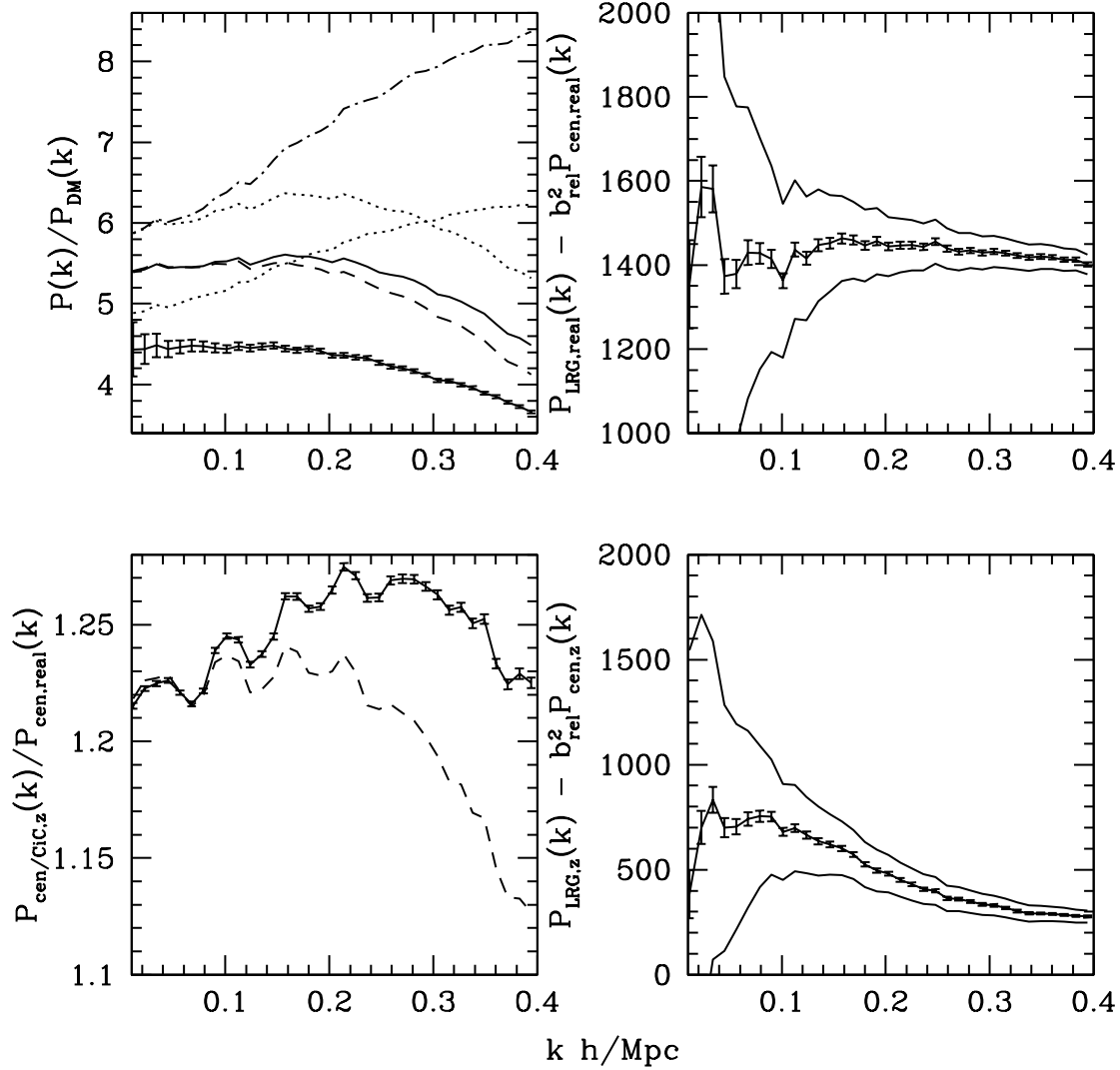


Fig. 7.— *Upper left:* The ratio of the monopole (i.e., angle-averaged) LRG $P(k)$ to the real space nonlinear dark matter spectrum for several MID subsamples: central LRGs only (solid curves); all LRGs (dotted curves); all LRGs after undergoing Tegmark et al. (2006) FOG compression (dash-dot curve); and CiC groups (dashed curve). For the first two subsamples we also show the real space $P(k)$, which have lower amplitude as $k \rightarrow 0$. *Upper right:* The real space $P(k)$ for the central and satellite LRGs minus b_{rel}^2 times $P_{\text{cen}}(k)$, the power spectrum of the central LRGs in real space, for $b_{\text{rel}} = 1.037, 1.042, 1.047$. *Lower left:* The solid curve shows the ratio of the central LRGs $P(k)$ in redshift space to the central LRGs $P(k)$ in real space, while the dashed curve shows the ratio for $P_{\text{CiC}}(k)$. *Lower right:* The redshift space monopole $P(k)$ for the central and satellite LRGs minus b_{rel}^2 times $P_{\text{cen}}(k)$, the power spectrum of the central LRGs in redshift space, for $b_{\text{rel}} = 1.037, 1.042, 1.047$. Error bars are similar for all curves in a figure, and so we show only one set for clarity.

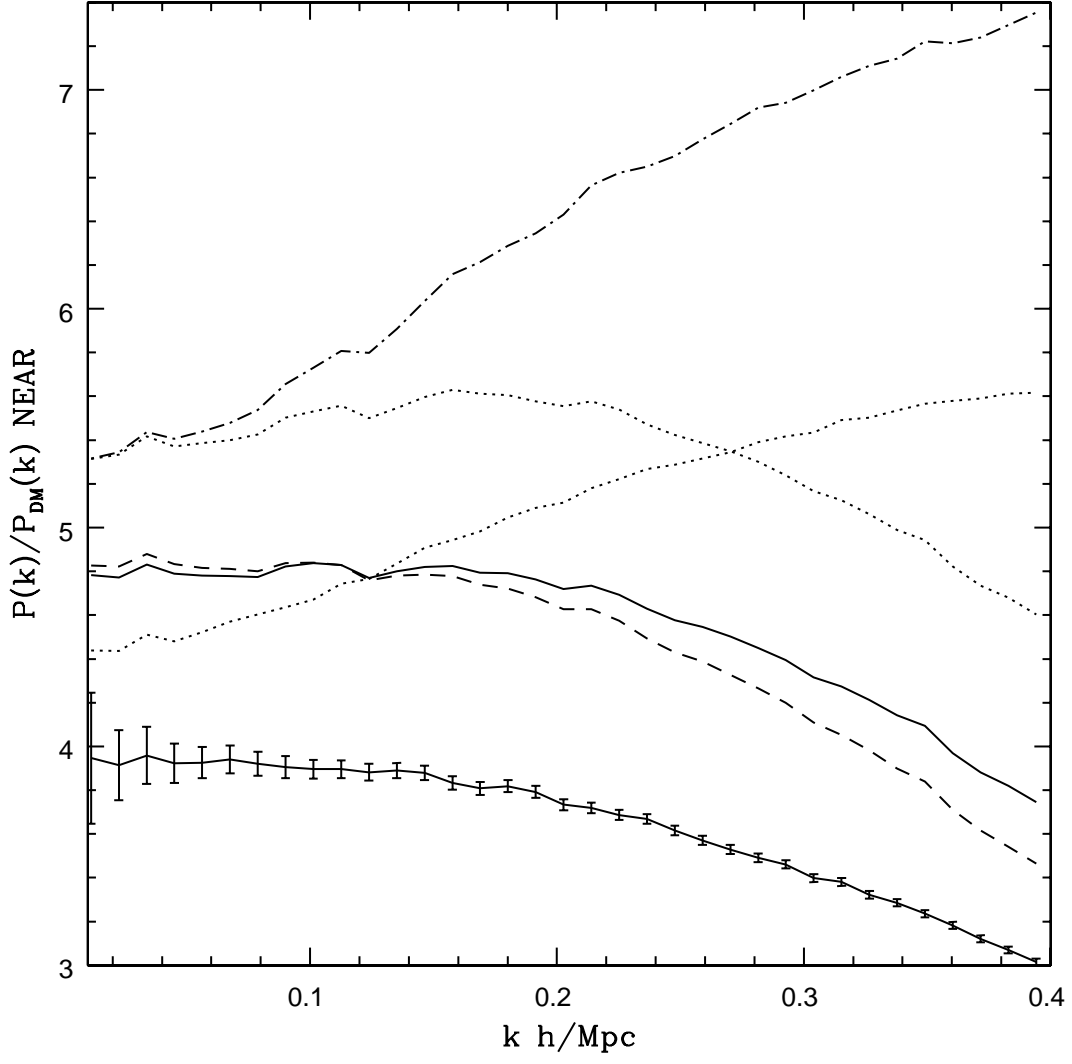


Fig. 8.— Same as Fig. 7 upper left panel, but for the NEAR sample.

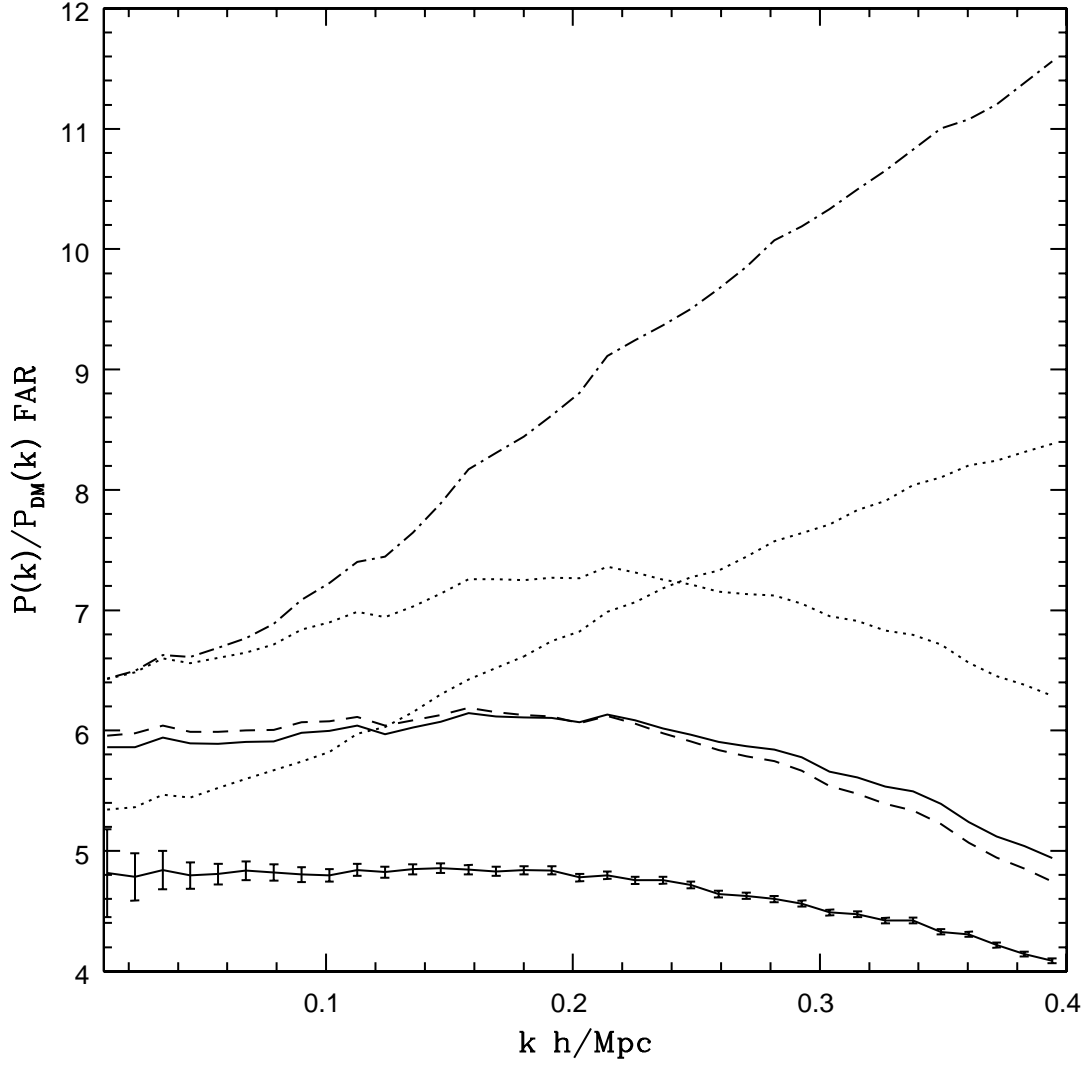


Fig. 9.— Same as Fig. 7 upper left panel, but for the FAR sample.

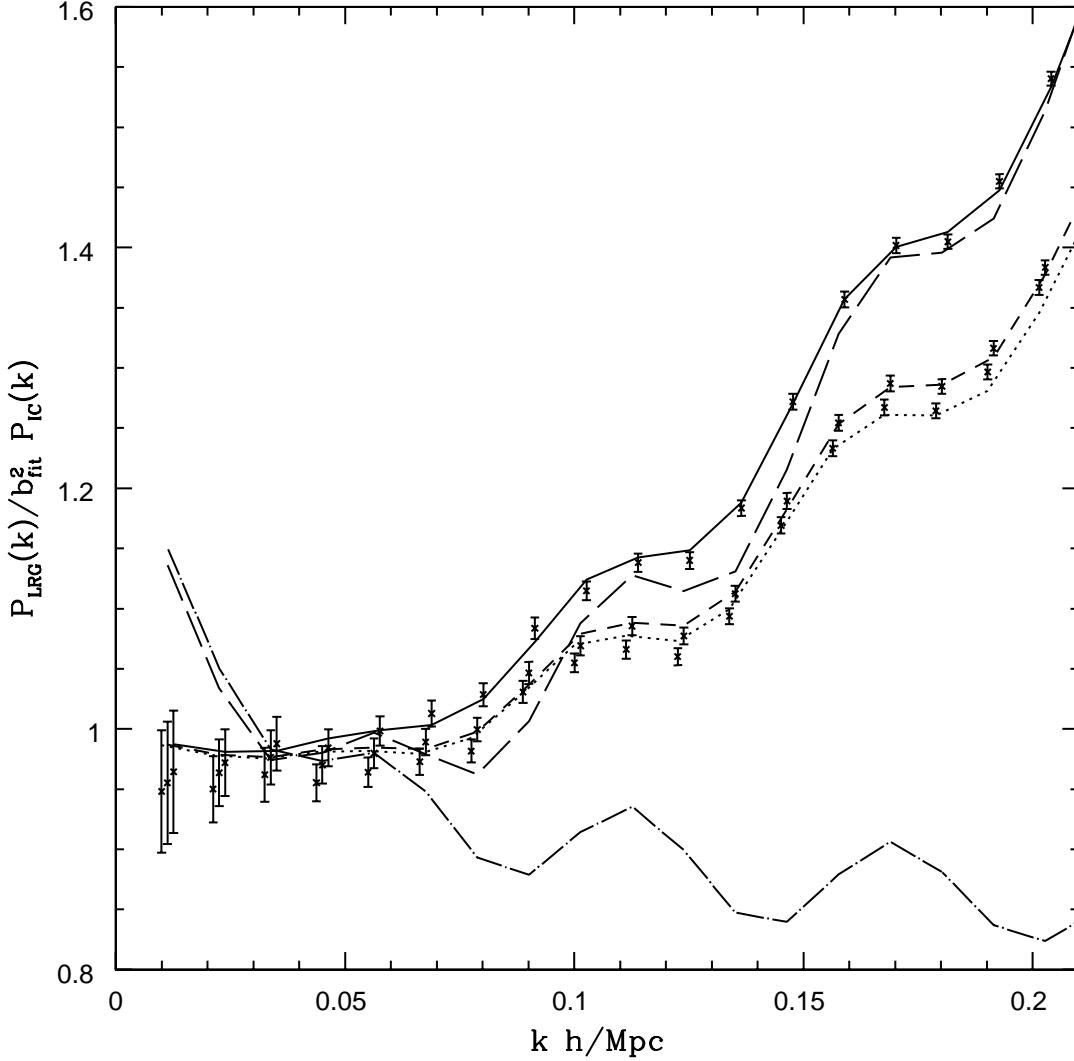


Fig. 10.— We separately fit our NEAR, MID, and FAR FOG-compressed samples to the nonlinear correction model of Tegmark et al. (2006) in Eqn. 6. Best fit parameters are $b_{NEAR}^2 = 5.61$, $Q_{NL,NEAR} = 17.4$ (dotted curve); $b_{MID}^2 = 6.15$, $Q_{NL,MID} = 18.4$ (dashed curve); and $b_{FAR}^2 = 6.67$, $Q_{NL,FAR} = 23.3$ (solid curve). Each subsample is scaled by the input power spectrum and the best fit b^2 value to accentuate the variation in the power spectrum shapes for these three samples. The NEAR and FAR data points and model curves have been shifted by $\pm 0.0013 h \text{ Mpc}^{-1}$ for clarity. The best fit Q_{NL} vanilla model in Tegmark et al. (2006), $A = 1.4$ and $Q_{NL} = 31$ in Eqn. 4, is shown by the long dashed curve. We have adjusted the normalization to agree at $k \sim 0.05$. The dot-dashed curve shows the ratio of the best fit linear power spectrum of Tegmark et al. (2006) and the linear power spectrum adopted in this work, demonstrating that Q_{NL} is degenerate with changes in the spectral shape, so that cosmological parameter information is lost to Q_{NL} .

$$C_{ij,tot} = \frac{1}{N_i} \left(P(k_i)P(k_j) + \frac{\alpha_{tot}}{\bar{n}} \right)^2 \delta_{ij}^K + \beta_{tot} B_{ij} \quad (32)$$

$$C_{ij,red} = \frac{1}{N_i} \left(P^2(k_i) + \frac{\alpha_{red}P(k_i)}{\bar{n}} + \frac{1}{\bar{n}^2 N_{mocks}} \right) \delta_{ij}^K + \beta_{red} B_{ij} \quad (33)$$

$$C_{ij,HOD} = \frac{1}{N_i} \left(\frac{\alpha_{HOD}P(k_i)}{\bar{n}} + \frac{N_{mocks}}{\bar{n}^2(N_{mocks} - 1)} \right) \delta_{ij}^K + \beta_{HOD} B_{ij}. \quad (34)$$

Note we use the measured $P(k_i)$ rather than the linear or model $P(k_i)$ in Eqns. 31-34. In the reduced covariance we expect the Poisson noise to be suppressed by $1/N_{mocks}$, the number of mock catalogs we produce for each TPM simulation. Eqn. 19 implies that $\alpha_{HOD} + \alpha_{red} = 2\alpha_{tot}$ for $\alpha_{tot} \approx 1$ and $\beta_{HOD} + \beta_{red} = \beta_{tot}$. Best fit values for each mock catalog type and each redshift subsample are listed in Table 4, where the fit included all modes with $k \leq 0.4$, and in Table 5, where the fit included all modes between $k = 0.05$ and $k = 0.2$. We present both in order to assess whether the non-Gaussian terms are growing with k . In Fig. 11 we show the measured and model C_{ij} values along the diagonal as well as three rows in the matrix for the MID reconstructed halo density field covariance matrix; the agreement with the model for other rows and redshift subsamples are similar.

We find $\alpha_{tot} \approx 1$ for all mocks and redshift subsamples, indicating that the standard shot noise contribution to the covariance is approximately correct for our mock catalogs. However, the relative distribution of this cross term between $C_{ij,HOD}$ and $C_{ij,red}$ varies; α_{HOD} is large for the FAR sample, where the number density is \sim half that in the MID and NEAR sample, and is also larger in the samples including satellites relative to the samples with only central objects. All three matrices have significant off-diagonal terms. β_{tot} is larger than unity for all samples, which may be expected since the LRGs are biased tracers, so the DC mode variance will increase by the factor b_{LRG}^2 . However, the amplitude does not scale with b^2 , but is much larger for the FAR sample. The best fit parameters do not change between real and redshift space for the same sample, but β_{tot} is larger for the samples including satellites. The reconstructed halo density field best fit parameters are consistent with the parameters of the sample of central galaxies, and the FOG-compressed sample is consistent with the sample of central and satellite galaxies. β_{tot} decreases slightly when fitting only to the $k = 0.05$ to $k = 0.2$ results, but the effect is small and so we adopt the parameters reported in Table 5 to carry out our covariance matrix calculations for the error bars on $P_{LRG}(k)$.

There is a modest decrease in the normalized diagonal covariance $C_{ii}/\bar{P}(k_i)^2$ ($\sim 5\%$) when satellite galaxies are included. However, this decrease is offset by the fact that the $P^2(k_i)$ has a larger nonlinear component, so not all of the available information will be about the linear component; furthermore, the large nonlinear correction introduces uncertainty in

extracting the linear component. Moreover, β_{tot} is larger for samples which include satellite galaxies, so the bandpowers will be more highly correlated for those mocks. We therefore conclude that the reduction in large scale bias by eliminating the satellite galaxies and reconstructing the halo density field is offset in the error budget by smaller off-diagonal covariance and smaller uncertainties in the nonlinear correction.

6.2.3. The Redshift Space Quadrupole

In linear theory the redshift space power spectrum can be decomposed into a monopole, quadrupole, and hexadecapole:

$$\beta = \frac{1}{b_{gal}} \frac{d \ln D}{d \ln a} \quad (35)$$

$$P_s(\mathbf{k}) = P(k) \left[\left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)L_0(\mu_{\mathbf{k}}) + \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right)L_2(\mu_{\mathbf{k}}) + \left(\frac{8}{35}\beta^2\right)L_4(\mu_{\mathbf{k}}) \right] \quad (36)$$

where L_i is the Legendre polynomial of order i , and $\mu_{\mathbf{k}} = \hat{\mathbf{k}} \cdot \hat{\mathbf{s}}$, with $\hat{\mathbf{s}}$ the direction along the LOS. The method of Tegmark et al. (2006) assumes Eqn. 36 with β independent of k to obtain the best estimate of the real space power spectrum from P_{gg} , P_{gv} , and P_{vv} . In Fig. 12 we examine the structure of the redshift space distortions in our mock catalogs. The first three panels show the quadrupole to monopole ratio for central LRGs (solid), the reconstructed halo density field (dashed), and the FOG-compressed density field (dash-dot). These are all similar and show a modest increase in the quadrupole to monopole ratio with k . This demonstrates that the FOG compression technique of Tegmark et al. (2006) successfully removes the effects of the FOGs induced by satellite galaxies. The oscillations probably result from the extra suppression of BAO features in redshift space (Eisenstein et al. 2007b). The central and satellite LRG sample (dotted) quadrupole to monopole ratio lies below the expected value at all k values, indicating that the FOG suppression of power is evident even in the linear regime. Therefore a sample where the effects of satellite FOGs have not been removed by either halo density field reconstruction or FOG compression will provide a biased estimate of β .

The bottom right panel shows the ratio of the redshift space monopole to the real space power spectrum for the sample of central galaxies only (solid) and central and satellite galaxies (dotted). The suppression of power by the satellite FOGs is also evident here. For central LRGs only, the ratio is consistent with the expected value from Eqn. 36, and between $k = 0$ and $k = 0.2$ this ratio varies by only $\sim 3\%$. While there is a clear scale dependence in the

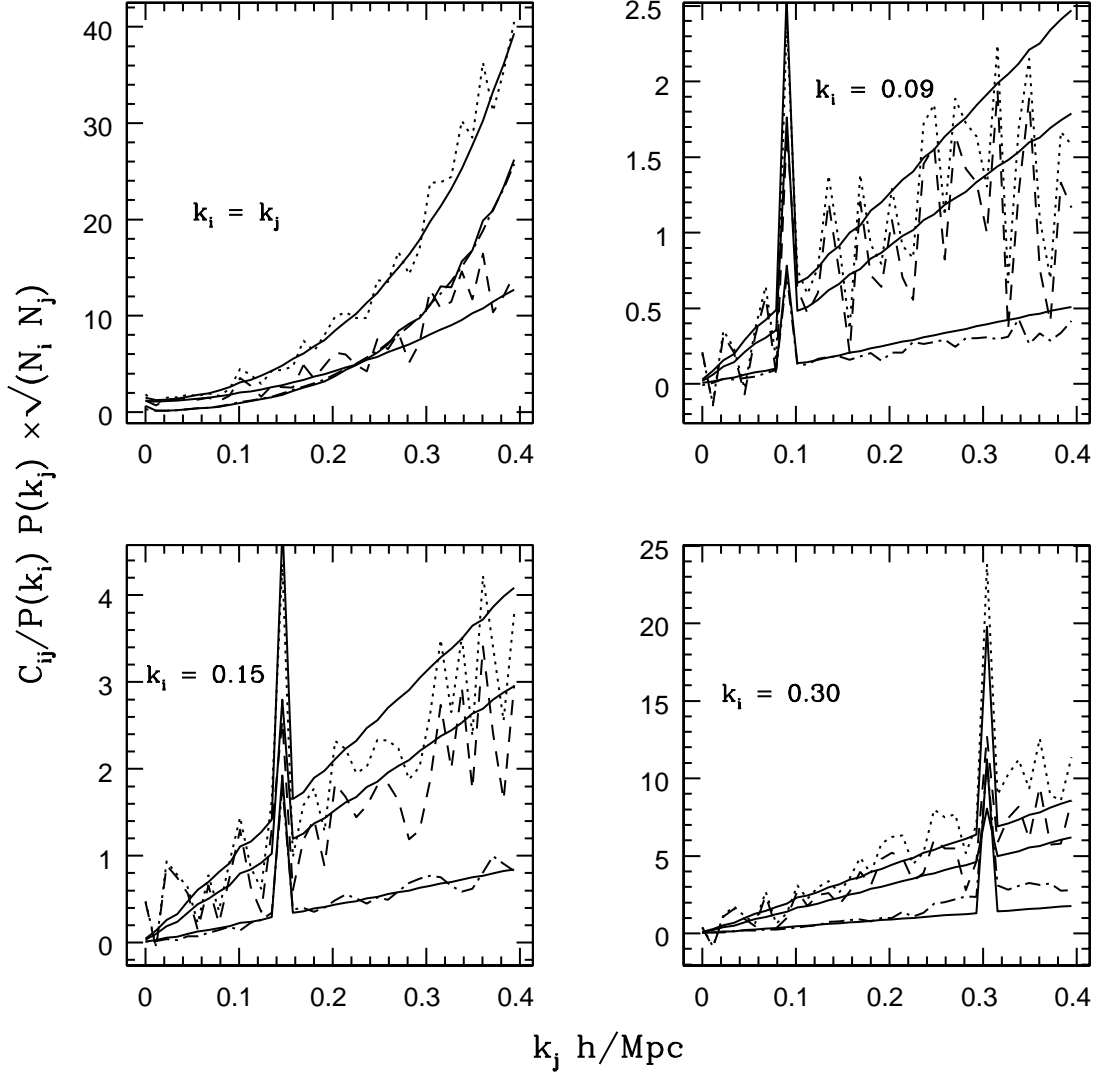


Fig. 11.— We plot the normalized covariance matrix elements $C_{ij}/P(k_i)P(k_j)$ scaled by $\sqrt{N_i N_j}$, the number modes in bins i and j . We plot our measurements of $C_{ij,tot}$ (dotted curve), $C_{ij,red}$ (dashed curve), and $C_{ij,HOD}$ (dot-dashed curve). The solid lines show the model fits from Table 4. The top left panel shows the diagonal elements, and the other panels show cross sections of the covariance matrix with $k_i = 0.09, 0.15$, and 0.3 h Mpc^{-1} .

quadrupole to monopole ratio for the central LRG sample, the large bias of LRGs means that the monopole redshift space power spectrum is nearly insensitive to the redshift space nonlinearities in the halo density field. The upper dotted curve shows $P_{s,sat}/P_{real,cen} \times b_{cen}^2/b_{sat}^2$. The increase in power in real space when the satellites are included is larger than the suppression of power in redshift space by their FOGs, so that the satellite monopole spectrum has more power at high k than the real space central LRGs.

7. Conclusions

This paper introduced and tested our algorithm for using the halo density field to estimate the underlying matter power spectrum. We found the nonlinear correction between this field and the underlying matter density field to be both smaller and more robust to variations in the effects of satellite galaxies than both the FOG-compressed density field used in the analysis of Tegmark et al. (2006) and the redshift space monopole power spectrum used in the analysis of Percival et al. (2007).

The parameters of our simulation set were selected to provide accurate two point halo density and velocity statistics to assure accurate representations of FOG features in our mock catalogs. To our knowledge, this is the first attempt to study the detailed effects of FOG treatment on the LRG power spectrum. We find that the FOG treatment can affect bandpowers even in the linear regime, $k \leq 0.1$. We first examined the nonlinear matter power spectrum of our 42 simulations. The ratio of the nonlinear to input matter power spectra is well described by a smearing of the BAOs as modeled by Eisenstein et al. (2007b) and a smooth increase in power with k that can be fit by a third order polynomial out to $k = 0.4 h \text{ Mpc}^{-1}$ or second order polynomial out to $k = 0.2 h \text{ Mpc}^{-1}$. We detect a substantial deviation from the predictions of halofit (Smith et al. 2003).

Fig. 7 demonstrates the main point of this work: satellite galaxies systematically alter the shape of the power spectrum *even* at $k < 0.1 h \text{ Mpc}^{-1}$. Extraction of cosmological information from the broadband shape of the power spectrum is already limited by systematics (Sanchez & Cole 2007) which we (and others) suggest can be attributed primarily to differences in the satellite contribution to the power spectrum. In this paper we demonstrate that while the FOG compression scheme in Tegmark et al. (2006) exacerbates these issues and requires a large nonlinear correction, the FOG features in the density field can be used to reconstruct the halo density field with high fidelity. The power spectrum of this field deviates

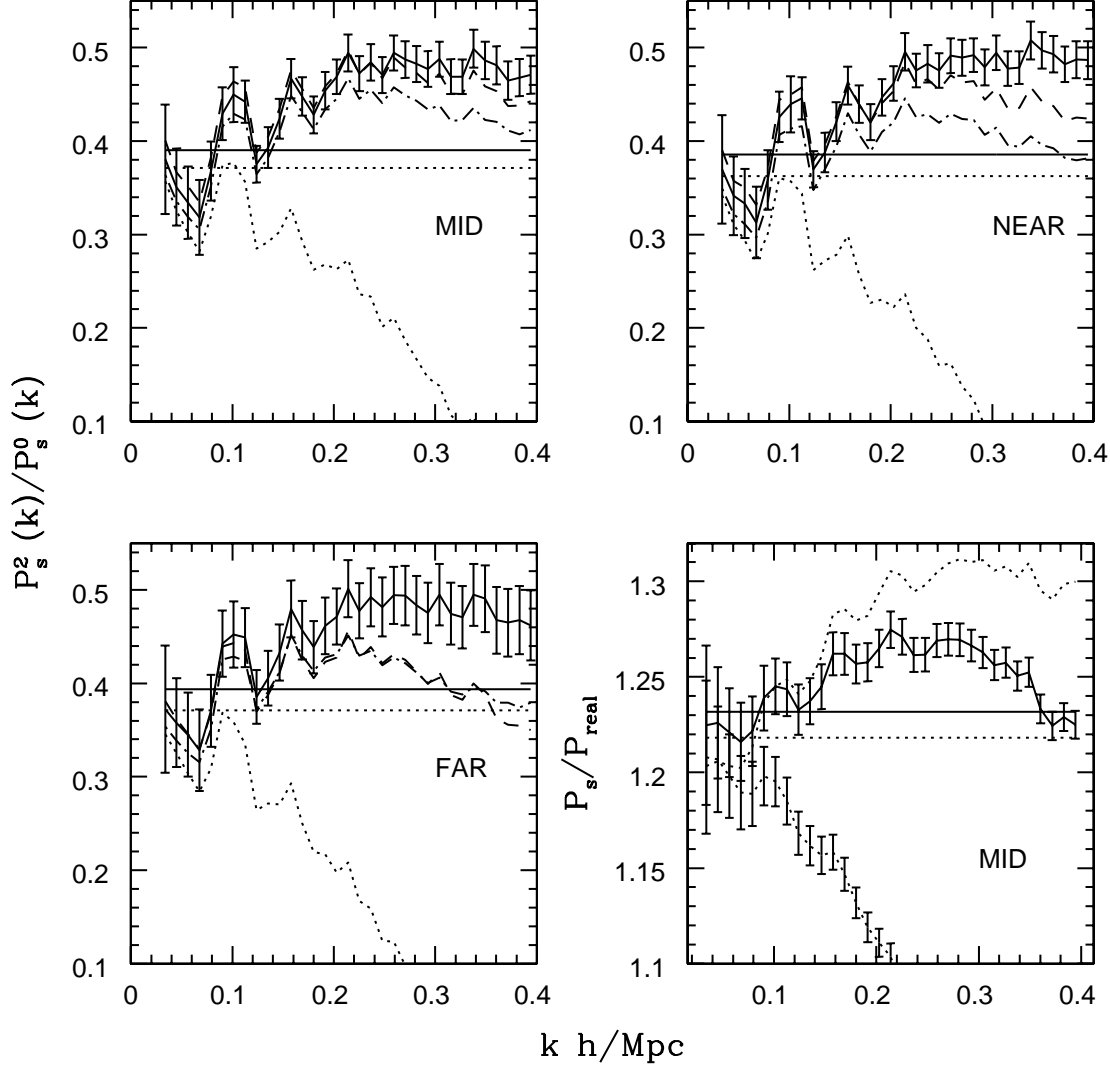


Fig. 12.— The bottom right panel shows the ratio of the redshift space monopole to the real space power spectrum for the MID sample. The straight lines show the prediction of Eqn. 36 for both samples. We fit the bias in Eqn. 35 to the large scale ratio P_{LRG}/P_{DM} . The solid curve is for the sample of central galaxies only ($b_{cen} = 2.11$) and the dotted curve with error bars includes both central and satellite galaxies ($b_{sat} = 2.23$). The upper dotted curve shows the ratio of the redshift space monopole of the central and satellite galaxies to the real space power spectrum of the central objects (scaled by b_{cen}^2/b_{sat}^2). The other three panels show the quadrupole to monopole ratio in redshift space. Error bars are only shown for the central galaxy sample (solid curve) for clarity. The central and satellite sample is shown by the dotted curve, the reconstructed halo density field as the dashed curve, and the FOG-compressed sample as the dash-dot curve. The expected values from Eqn. 36 are shown as the straight lines for the central (solid) and central and satellite (dotted) samples.

from the dark matter power spectrum at the $\leq 4\%$ level for $k \leq 0.2 h \text{ Mpc}^{-1}$ and $\leq 1\%$ level for $k \leq 0.1 h \text{ Mpc}^{-1}$, where cosmological analyses usually restrict themselves. Moreover, we have shown that this correction changes only slightly between the NEAR, MID, and FAR LRG reconstructed halo density fields, while the FOG compressed mocks have a much larger variation between samples. Therefore, we can hope to push cosmological analyses to larger k using the reconstructed halo density field as a tracer of the underlying matter density field fluctuations— particularly for galaxy samples like the LRGs which are spread over a large redshift range and are not volume-limited, and thus have substantial variation in the satellite contribution to the power spectrum with redshift.

While we have not addressed the variation of the nonlinear correction to the reconstructed halo density field as a function of cosmological parameters, we have designed the form of our correction to minimize the variation with cosmology. Other researchers (e.g., Habib et al. 2007) are studying the dark matter power spectrum as a function of cosmology. We expect that our nonlinear correction P_{LRG}/P_{DM} will remain small (of order 4% below $k = 0.2 h \text{ Mpc}^{-1}$) as the cosmology is varied, so the variation of this small correction should be even smaller. Therefore, instead of introducing a nuisance parameter for the nonlinear correction, we propose that the amplitude of the correction should be taken as the error on its value in cosmological parameter analyses, or relatively strong priors on the amplitude of the correction to the reconstructed halo density field be introduced.

In this work we have also investigated the properties of the power spectrum covariance matrix for the dark matter, as well as the mock galaxy catalogs divided into 6 different samples: central galaxies in real and redshift space, central and satellite galaxies in real and redshift space, our reconstructed halo density field in redshift space, and the Tegmark et al. (2006) FOG compressed galaxy density field in redshift space. All of these samples were well-modeled by a diagonal matrix with the usual Gaussian and Poisson shot noise terms plus the beat-coupling term presented in Hamilton et al. (2006). We expect that this will be a useful model for fitting the survey covariance matrix, where $P(k_b)$ is replaced by $P(k_{survey})$ with k_{survey} determined by some effective survey size.

Finally we examined the structure of the redshift space distortions as a function of k using the quadrupole. Both the reconstructed halo density field and the FOG-compressed mock LRG samples reproduce the modest k dependence of the halo density field quadrupole to monopole ratio. Since the LRGs are so highly biased, this scale dependence causes a $\lesssim 3\%$ deviation in the redshift space monopole to real space power spectrum ratio out to $k = 0.2$.

When satellites are included in the sample without FOG compression, the quadrupole to monopole ratio is lower than the linear value for the entire k range accessible in our simulations and will significantly bias the estimate of β .

8. Acknowledgments

We thank Daniel Eisenstein for providing our SDSS LRG sample and Michael Blanton for providing the inverse random catalogs and Tycho2 catalog. We thank Jeremy Tinker for providing the SO halo-finding code used to produce our halo catalogs. We thank Will Percival and Licia Verde for excellent discussions.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

Computer simulations were supported by the National Science Foundation through TeraGrid resources provided by Pittsburgh Supercomputing Center and the National Center for Supercomputing Applications under grant AST070021N; simulations were also performed at the TIGRESS high performance computer center at Princeton University, which is jointly supported by the Princeton Institute for Computational Science and Engineering and the Princeton University Office of Information Technology.

This work draws on Chapters 3 and 4 of BAR’s Ph.D. thesis in Princeton University Department of Physics. BAR gratefully acknowledges support from both the National Science Foundation Graduate Research Fellowship and PIRE program. This project was supported

by NSF Grant 0707731. DNS thanks the APC (Universite de Paris VII) for its hospitality while this work was completed.

REFERENCES

- Baugh, C. M., & Efstathiou, G. 1994, *MNRAS*, 270, 183
- Bode, P., & Ostriker, J. P. 2003, *ApJS*, 145, 1
- Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propriis, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., & Taylor, K. 2005, *MNRAS*, 362, 505
- Cooray, A., & Sheth, R. 2002, *Phys. Rep.*, 372, 1
- Crocce, M., & Scoccimarro, R. 2008, *Phys. Rev. D*, 77, 023533
- Dunkley, J., Komatsu, E., Nolte, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. 2008, *ArXiv e-prints*, 803
- Eisenstein, D. J., Annis, J., Gunn, J. E., Szalay, A. S., Connolly, A. J., Nichol, R. C., Bahcall, N. A., Bernardi, M., Burles, S., Castander, F. J., Fukugita, M., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lupton, R. H., Narayanan, V., Postman, M., Reichart, D. E., Richmond, M., Schneider, D. P., Schlegel, D. J., Strauss, M. A., SubbaRao, M., Tucker, D. L., Vanden Berk, D., Vogeley, M. S., Weinberg, D. H., & Yanny, B. 2001, *AJ*, 122, 2267
- Eisenstein, D. J., & Hu, W. 1998, *ApJ*, 496, 605
- Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007a, *ApJ*, 664, 675
- Eisenstein, D. J., Seo, H.-J., & White, M. 2007b, *ApJ*, 664, 660
- Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, *ApJ*, 426, 23
- Frenk, C. S., White, S. D. M., Davis, M., & Efstathiou, G. 1988, *ApJ*, 327, 507

- Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., & Williams, B. 2007, *Phys. Rev. D*, 76, 083503
- Hamann, J., Hannestad, S., Melchiorri, A., & Wong, Y. Y. Y. 2008, *Journal of Cosmology and Astro-Particle Physics*, 7, 17
- Hamilton, A. J. S., Rimes, C. D., & Scoccimarro, R. 2006, *MNRAS*, 371, 1188
- Hockney, R. W., & Eastwood, J. W. 1981, *Computer Simulation Using Particles* (Computer Simulation Using Particles, New York: McGraw-Hill, 1981)
- Huff, E., Schulz, A. E., White, M., Schlegel, D. J., & Warren, M. S. 2007, *Astroparticle Physics*, 26, 351
- Jing, Y. P. 2005, *ApJ*, 620, 559
- Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2008, *ArXiv e-prints*, 803
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, *ApJ*, 609, 35
- Lokas, E. L., Juszkievicz, R., Bouchet, F. R., & Hivon, E. 1996, *ApJ*, 467, 1
- Matsumoto, M., & Nishimura, T. 1998, *ACM Transactions on Modeling and Computer Simulation*, 8, 3
- Neyrinck, M. C., Szapudi, I., & Rimes, C. D. 2006, *MNRAS*, 370, L66
- Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Percival, W. J., Nichol, R. C., Eisenstein, D. J., Frieman, J. A., Fukugita, M., Loveday, J., Pope, A. C., Schneider, D. P., Szalay, A. S., Tegmark, M., Vogeley, M. S., Weinberg, D. H., Zehavi, I., Bahcall, N. A., Brinkmann, J., Connolly, A. J., & Meiksin, A. 2007, *ApJ*, 657, 645
- Reid, B. A. 2008, PhD thesis, Princeton University
- Reid, B. A., & Spergel, D. N. 2008, *ArXiv e-prints*
- Sanchez, A. G., & Cole, S. 2007, *ArXiv e-prints*, 708

- Schulz, A. E., & White, M. 2006, *Astroparticle Physics*, 25, 172
- Seljak, U. 2000, *MNRAS*, 318, 203
- Sirko, E. 2005, *ApJ*, 634, 728
- Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., & Couchman, H. M. P. 2003, *MNRAS*, 341, 1311
- Tegmark, M., Blanton, M. R., Strauss, M. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., Weinberg, D. H., Zehavi, I., Berlind, A., Budavari, T., Connolly, A., Eisenstein, D. J., Finkbeiner, D., Frieman, J. A., Gunn, J. E., Hamilton, A. J. S., Hui, L., Jain, B., Johnston, D., Kent, S., Lin, H., Nakajima, R., Nichol, R. C., Ostriker, J. P., Pope, A., Scranton, R., Seljak, U., Sheth, R. K., Stebbins, A., Szalay, A. S., Szapudi, I., Verde, L., Xu, Y., Annis, J., Bahcall, N. A., Brinkmann, J., Burles, S., Castander, F. J., Csabai, I., Loveday, J., Doi, M., Fukugita, M., Gott, J. R. I., Hennessy, G., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., Lee, B. C., Lupton, R. H., McKay, T. A., Kunszt, P., Munn, J. A., O’Connell, L., Peoples, J., Pier, J. R., Richmond, M., Rockosi, C., Schneider, D. P., Stoughton, C., Tucker, D. L., Vanden Berk, D. E., Yanny, B., & York, D. G. 2004, *ApJ*, 606, 702
- Tegmark, M., Eisenstein, D. J., Strauss, M. A., Weinberg, D. H., Blanton, M. R., Frieman, J. A., Fukugita, M., Gunn, J. E., Hamilton, A. J. S., Knapp, G. R., Nichol, R. C., Ostriker, J. P., Padmanabhan, N., Percival, W. J., Schlegel, D. J., Schneider, D. P., Scoccimarro, R., Seljak, U., Seo, H.-J., Swanson, M., Szalay, A. S., Vogeley, M. S., Yoo, J., Zehavi, I., Abazajian, K., Anderson, S. F., Annis, J., Bahcall, N. A., Bassett, B., Berlind, A., Brinkmann, J., Budavari, T., Castander, F., Connolly, A., Csabai, I., Doi, M., Finkbeiner, D. P., Gillespie, B., Glazebrook, K., Hennessy, G. S., Hogg, D. W., Ivezić, Ž., Jain, B., Johnston, D., Kent, S., Lamb, D. Q., Lee, B. C., Lin, H., Loveday, J., Lupton, R. H., Munn, J. A., Pan, K., Park, C., Peoples, J., Pier, J. R., Pope, A., Richmond, M., Rockosi, C., Scranton, R., Sheth, R. K., Stebbins, A., Stoughton, C., Szapudi, I., Tucker, D. L., Berk, D. E. V., Yanny, B., & York, D. G. 2006, *Phys. Rev. D*, 74, 123507
- Tinker, J. L. 2007, *MNRAS*, 374, 477
- Tinker, J. L., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M. S., Yepes, G., Gottlober, S., & Holz, D. E. 2008, *ArXiv e-prints*, 803
- Vale, A., & Ostriker, J. P. 2006, *MNRAS*, 371, 1173
- Verde, L., & Peiris, H. 2008, *Journal of Cosmology and Astro-Particle Physics*, 7, 9

- Yoo, J., Weinberg, D. H., Tinker, J. L., Zheng, Z., & Warren, M. S. 2008, ArXiv e-prints
- Zehavi, I., Eisenstein, D. J., Nichol, R. C., Blanton, M. R., Hogg, D. W., Brinkmann, J., Loveday, J., Meiksin, A., Schneider, D. P., & Tegmark, M. 2005, ApJ, 621, 22

9. Appendix A: Resolution Study Results

The goal of the resolution study detailed in Chapter 3 of Reid (2008) was to determine the minimum N -body simulation mass resolution satisfying two criteria.

- The minimum halo mass is small enough to accommodate broad $N_{cen}(M)$ functions (i.e., large values of $\sigma_{\log M}$). The width of this function is dictated by the observed large-scale clustering amplitude, either in $P_{LRG}(k)$ or $w_p(r_p)$, which determines $\sigma_8 b_{LRG}$.
- The finite mass resolution does not systematically bias the small scale n -point halo statistics that must be accurate in order to test our method of halo density field reconstruction. We argue in Reid (2008) that agreement of several two-point clustering and velocity statistics and the CiC group multiplicity functions is sufficient for accurate mock FOG structures. Furthermore, the small observed deviations in the halo mass function at low M can be absorbed by the parameters of $N_{cen}(M)$.

For the cosmological parameters of our mock catalogs and the observed LRG clustering strength, the first condition places the more stringent requirement on the simulation mass resolution. To establish the level of systematics relevant to the second condition, we compare the results of one N -body simulation with our adopted parameters $L_{box} = 558 h^{-1}$ Mpc and $N_{p,med} = 512^3$ ($M_{p,med} = 1.43 \times 10^{11} M_\odot$) with a higher resolution simulation ($N_{p,high} = 640^3$, $M_{p,high} = 7.33 \times 10^{10} M_\odot$) and a lower resolution simulation ($N_{p,low} = 384^3$, $M_{p,low} = 3.39 \times 10^{11} M_\odot$) of the same initial conditions. We highlight the findings of this comparison below for the simulation outputs near our $z_{MID} = 0.342$, and refer the interested reader to Reid (2008) for a more thorough discussion.

9.1. Resolution Effects in the Halo Mass Function and the Halo Occupation Distribution

The mass functions from the low, medium, and high resolution simulations are in excellent agreement with one another and within a few percent of the Tinker et al. (2008) mass function. There is a slight increase ($\sim 5\%$) in the halo mass function as the aggressive mass limit $50M_p$ of our SO halo catalogs is reached; this should be kept in mind in future analyses, but is not inherently problematic since slight changes in the mass function in this regime are degenerate with small variations in the form of $N_{cen}(M)$.

We produce LRG mock catalogs at $\sigma_{logM} = 0.6$, the value that matches the observed LRG clustering amplitude, in the low, medium, and high resolution simulations. For the high resolution simulation, only 2.6% of the mock LRGs occupy halos below the halo catalog mass limit of the medium resolution simulation, and only 11% are below $100M_{p,med}$. Therefore, the abrupt cutoff in $N_{cen}(M)$ induced by our halo catalog mass limit will show minimal differences with a mock catalog derived from a halo catalog extending to a substantially lower mass limit. In addition, even if there are slight systematics in the pair statistics of a halo catalog in the $\sim 50 - 100M_{p,med}$ mass range, they will be suppressed by at least a factor of ~ 10 .

9.2. Two-point clustering and velocity statistics

There is a slight but systematic decrease in power with k as the simulation resolution decreases. This is a 0.2% effect at $k = 0.2 h \text{ Mpc}^{-1}$ between the medium and high resolution simulations, well within the statistical errors and modeling uncertainties between the LRGs and dark matter density fields.

We split the halos into two mass bins: $50 - 99M_{p,med}$ and $M \geq 100M_{p,med}$. This corresponds to 98 – 195 particles per halo for the low mass bin of the high resolution simulation. We did not detect any significant resolution-dependent differences in $P(k)$ and $\xi(r)$, other than a $\lesssim 1\%$ change in the large scale bias of the halo samples with respect to the matter. Fig. 13 shows that the difference between the medium and high resolution simulation $\xi_{ij}(r)$, for both auto- and cross-correlations of the low and high mass halo bins, is consistent with 0 at small scales ($r \lesssim 15 h^{-1} \text{ Mpc}$); this is true out to $150 h^{-1} \text{ Mpc}$. Fig. 14 shows the velocity correlation statistic $\langle \mathbf{v}_i(\mathbf{r}) \cdot \mathbf{v}_j(\mathbf{r}') \rangle / \sqrt{\langle \mathbf{v}_i^2 \rangle \langle \mathbf{v}_j^2 \rangle}$ for halo mass bins i and j . We find $\leq 1\%$ level agreement between the medium and high resolution simulations.

In Fig. 14 we also examine the relative halo velocities along the vector separating the halos: $\langle (\mathbf{v}_i(\mathbf{r}) - \mathbf{v}_j(\mathbf{r}')) \cdot (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'| \rangle$ and find no evidence for systematic bias between the medium and high resolution simulations.

9.3. Mock catalog statistics: two point clustering statistics and multiplicity bias $b(n_{sat})$

In Reid (2008) we compare both the power spectrum $P(k)$ and the projected correlation function $w_p(r_p)$ of the medium and high resolution simulations, for which we find systematic deviations smaller than $\sim 1/6$ the statistical error induced by the stochasticity of the halo occupation. We quantify the difference between the true one-halo group density field and the CiC group density field by the multiplicity bias $b(n_{sat}) = N_{CiC}(n_{sat})/N_{true}(n_{sat})$. We show that $b(n_{sat})$ systematically varies by less than the amount of variation induced by the halo occupation stochasticity for mock catalogs with $\sigma_{logM} = 0.8$, which pushes to lower mass limits than our better matched $\sigma_{logM} = 0.6$ catalogs.

We conclude that the selected mass resolution of our simulations is well-optimized to produce a large simulation volume over several realizations and with well-controlled systematic biases induced by the finite mass resolution.

10. Appendix B: Power Spectra and Covariance Matrix Best Fit Parameters

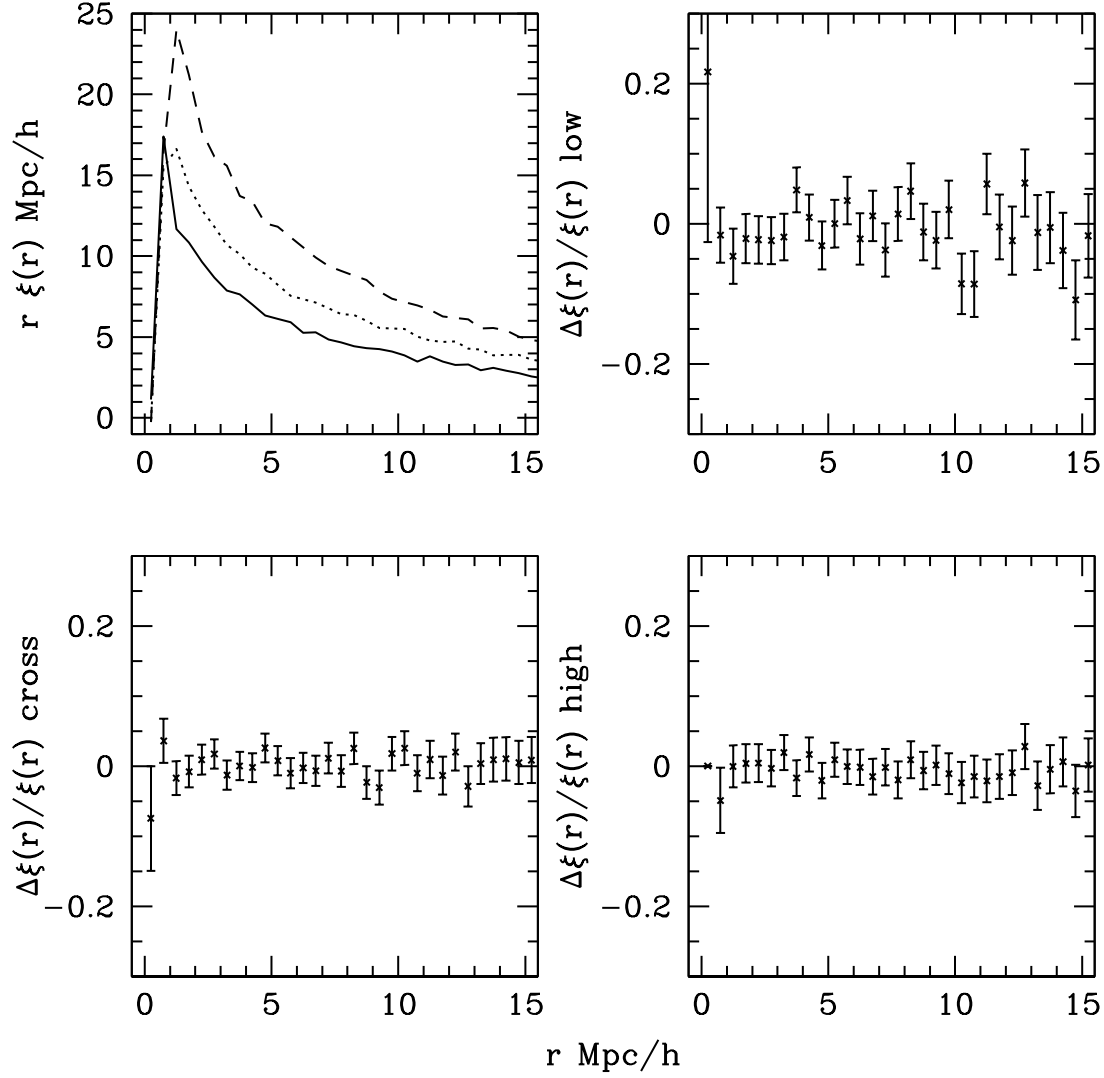


Fig. 13.— *Upper left:* $r\xi_{ij}(r)$ on small scales for the low mass halos auto-correlation (solid), high mass halos auto-correlation (dashed), and their cross correlation (dotted). Also shown are $(\xi_{high}(r) - \xi_{med}(r))/\xi_{high}(r)$ computed in bins of size $\Delta r = 0.5 h^{-1}$ Mpc for the low mass halos auto-correlation (upper right), high mass halos auto-correlation (lower right), and their cross correlation (lower left). The agreement for each correlation extends to $r = 150 h^{-1}$ Mpc (not shown).

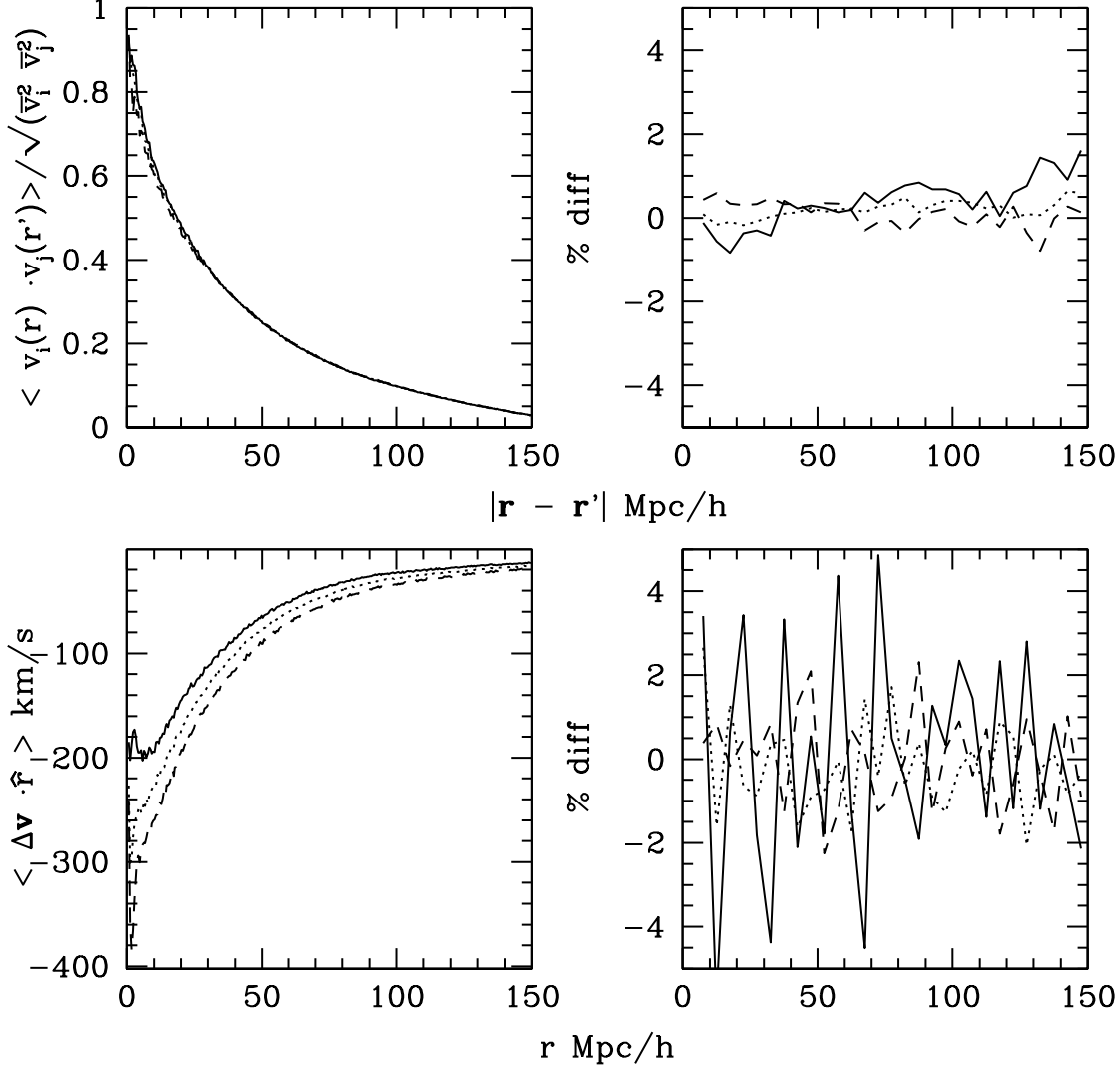


Fig. 14.— *Upper left:* The normalized halo-halo velocity correlation function, $\langle \mathbf{v}_i(\mathbf{r}) \cdot \mathbf{v}_j(\mathbf{r}') \rangle / \sqrt{\langle \mathbf{v}_i^2 \rangle \langle \mathbf{v}_j^2 \rangle}$, vs $|\mathbf{r} - \mathbf{r}'|$ for the low mass halos auto-correlation (solid), high mass halos auto-correlation (dashed), and their cross correlation (dotted) binned in intervals of $0.5 h^{-1} \text{ Mpc}$ in $|\mathbf{r} - \mathbf{r}'|$. *Upper right:* The fractional difference in the normalized halo-halo velocity correlation between the high and medium resolution simulations binned in intervals of $5 h^{-1} \text{ Mpc}$ in $|\mathbf{r} - \mathbf{r}'|$. *Lower panels:* Same as the upper panels, but for the halo-halo relative velocity correlation function along the halo separation vector, $\langle (\mathbf{v}_i(\mathbf{r}) - \mathbf{v}_j(\mathbf{r}')) \cdot (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'| \rangle$.

Table 3. Fits to nonlinear power spectra. $k_{max,fit}$ and k_{BAO} are in units of $h \text{ Mpc}^{-1}$. For $P_{DM}(k)/P_{IC}(k)$ we fit to Eqn. 30. When $k_{max,fit} = 0.2$ we hold $a_3 = 0$. Some of the redshift space mock catalog fits are not well-behaved at very small k 's, and the corrections to a constant should be suppressed at $k \lesssim 0.05$. This also makes a_0 deviate from the large scale bias, given in the last column and fit using only $k < 0.06$. $P_{sat,z}(k)$ denotes the power spectrum of a catalog containing central and satellite galaxies in redshift space, $P_{CiC,z}(k)$ denotes the power spectrum of the reconstructed halo density field in redshift space, and $P_{teg,z}(k)$ denotes an FOG compressed following Tegmark et al. (2006).

sample	function	$k_{max,fit}$	k_{BAO}	a_0	a_1	a_2	a_3	b_{LRG}^2
NEAR	$P_{DM}(k)/P_{IC}(k)$	0.2	0.12	1.025	-1.109	9.296	–	–
NEAR	$P_{DM}(k)/P_{IC}(k)$	0.4	0.13	1.027	-1.305	12.05	-9.136	–
NEAR	$P_{sat,z}(k)/P_{DM}(k)$	0.2	–	5.14	5.38	-15.96	–	5.38
NEAR	$P_{sat,z}(k)/P_{DM}(k)$	0.4	–	5.01	7.91	-28.72	15.08	5.38
NEAR	$P_{CiC,z}(k)/P_{DM}(k)$	0.2	–	4.74	2.11	-12.80	–	4.83
NEAR	$P_{CiC,z}(k)/P_{DM}(k)$	0.4	–	4.66	3.68	-20.45	8.40	4.83
NEAR	$P_{teg,z}(k)/P_{DM}(k)$	0.2	–	5.15	4.85	7.78	–	5.42
NEAR	$P_{teg,z}(k)/P_{DM}(k)$	0.4	–	5.03	6.44	6.41	-20.33	5.42
MID	$P_{DM}(k)/P_{IC}(k)$	0.2	0.13	1.025	-1.042	8.614	–	–
MID	$P_{DM}(k)/P_{IC}(k)$	0.4	0.14	1.026	-1.199	11.06	-8.426	–
MID	$P_{sat,z}(k)/P_{DM}(k)$	0.2	–	5.67	6.89	-17.9	–	5.99
MID	$P_{sat,z}(k)/P_{DM}(k)$	0.4	–	5.54	9.21	-28.3	8.54	5.99
MID	$P_{CiC}(k)/P_{DM}(k)$	0.2	–	5.30	3.43	-14.7	–	5.46
MID	$P_{CiC}(k)/P_{DM}(k)$	0.4	–	5.21	5.13	-21.8	4.35	5.46
MID	$P_{teg}(k)/P_{DM}(k)$	0.2	–	5.68	6.11	8.23	–	6.03
MID	$P_{teg}(k)/P_{DM}(k)$	0.4	–	5.58	7.27	10.4	-28.4	6.03
FAR	$P_{DM}(k)/P_{IC}(k)$	0.2	0.13	1.025	-0.994	8.144	–	–
FAR	$P_{DM}(k)/P_{IC}(k)$	0.4	0.15	1.025	-1.126	10.39	-7.945	–
FAR	$P_{sat,z}(k)/P_{DM}(k)$	0.2	–	6.12	9.72	-19.26	–	6.58
FAR	$P_{sat,z}(k)/P_{DM}(k)$	0.4	–	5.89	13.96	-37.46	11.39	6.58
FAR	$P_{CiC}(k)/P_{DM}(k)$	0.2	–	5.79	4.36	-13.97	–	6.00
FAR	$P_{CiC}(k)/P_{DM}(k)$	0.4	–	5.64	6.94	-23.48	–	6.00
FAR	$P_{teg}(k)/P_{DM}(k)$	0.2	–	6.16	7.90	26.43	–	6.63
FAR	$P_{teg}(k)/P_{DM}(k)$	0.4	–	5.96	10.49	26.82	-45.03	6.63

Table 4. Fits for the mock catalog covariance matrices using Eqns. 31 - 34; all bandpowers between $k = 0.01$ and $k = 0.4$ are included in the fits. ‘cen, real/redshift’ denotes a sample of only central LRGs in real/redshift space; ‘sat, real/redshift’ denotes a sample with both central and satellite galaxies in real/redshift space; ‘CiC, redshift’ denotes the reconstructed halo density field using the CiC method, and ‘Teg, redshift’ denotes the FOG-compressed sample according to Tegmark et al. (2006).

Catalog	α_{HOD}	α_{red}	$2\alpha_{tot}$	β_{HOD}	β_{red}	β_{tot}
NEAR cen, real	1.07	1.10	2.00	0.26	0.94	1.26
NEAR sat, real	1.30	0.89	1.98	0.36	1.58	1.97
NEAR cen, redshift	1.11	0.99	2.02	0.26	0.95	1.29
NEAR sat, redshift	1.27	0.63	2.01	0.29	1.28	1.64
NEAR CiC, redshift	1.11	0.92	2.02	0.26	0.94	1.30
NEAR Teg, redshift	1.39	0.74	1.98	0.35	1.60	1.99
MID cen, real	1.14	1.0	2.00	0.34	1.21	1.62
MID sat, real	1.32	0.77	1.99	0.42	1.71	2.20
MID cen, redshift	1.17	0.93	2.01	0.34	1.19	1.63
MID sat, redshift	1.30	0.61	2.00	0.36	1.47	1.92
MID CiC, redshift	1.16	0.86	2.01	0.34	1.18	1.63
MID Teg, redshift	1.41	0.58	1.97	0.41	1.73	2.20
FAR cen, real	1.65	0.50	2.05	1.32	1.62	3.13
FAR sat, real	1.85	0.38	2.07	1.71	2.67	4.50
FAR cen, redshift	1.66	0.42	2.05	1.25	1.64	3.13
FAR sat, redshift	1.79	0.16	2.07	1.37	2.08	3.70
FAR CiC, redshift	1.66	0.37	2.06	1.25	1.69	3.21
FAR Teg, redshift	1.90	0.13	2.06	1.54	2.53	4.19

Table 5. Fits for the mock catalog covariance matrices using Eqns. 31 - 34; all bandpowers between $k = 0.05$ and $k = 0.2$ are included in the fits. Samples are the same as in Table 4.

Catalog	α_{HOD}	α_{red}	$2\alpha_{tot}$	β_{HOD}	β_{red}	β_{tot}
NEAR cen, real	1.05	0.85	1.96	0.23	0.86	1.14
NEAR sat, real	1.22	0.62	1.94	0.29	1.43	1.76
NEAR cen, redshift	1.07	0.71	1.90	0.23	0.94	1.23
NEAR sat, redshift	1.22	0.51	1.86	0.26	1.33	1.66
NEAR CiC, redshift	1.07	0.68	1.90	0.24	0.97	1.27
NEAR Teg, redshift	1.28	0.45	1.86	0.29	1.47	1.82
MID cen, real	1.08	0.84	1.96	0.28	1.14	1.49
MID sat, real	1.21	0.62	1.93	0.35	1.60	2.02
MID cen, redshift	1.11	0.74	1.93	0.30	1.20	1.58
MID sat, redshift	1.23	0.54	1.88	0.33	1.55	1.96
MID CiC, redshift	1.10	0.72	1.91	0.29	1.24	1.61
MID Teg, redshift	1.29	0.46	1.88	0.36	1.66	2.10
FAR cen, real	1.57	0.36	2.00	1.08	1.66	2.87
FAR sat, real	1.68	0.21	2.00	1.35	2.51	3.98
FAR cen, redshift	1.57	0.26	1.99	1.06	1.70	2.94
FAR sat, redshift	1.67	0.13	1.98	1.20	2.28	3.65
FAR CiC, redshift	1.58	0.26	2.00	1.06	1.84	3.09
FAR Teg, redshift	1.73	0.06	1.99	1.31	2.47	3.93